

## BAB 3 METODOLOGI PENELITIAN

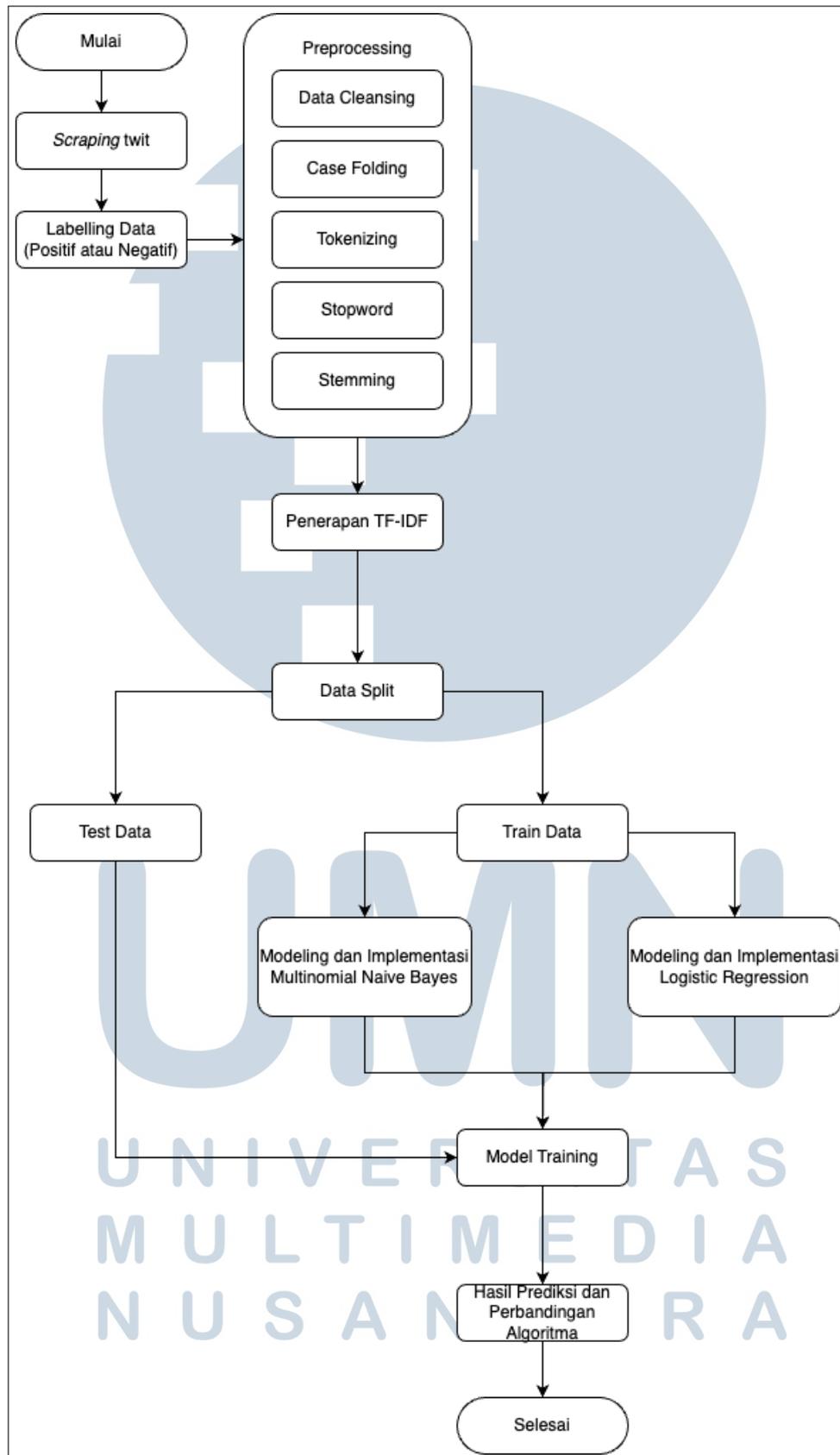
### 3.1 Gambaran Umum Objek Penelitian

ChatGPT merupakan program *artificial intelligence* (AI) atau kecerdasan buatan yang memungkinkan pengguna melakukan percakapan interaktif dengan AI tersebut. Selain mampu melakukan percakapan interaktif dengan manusia, ChatGPT mampu memberikan informasi spesifik yang diinginkan manusia maupun menyelesaikan tugas-tugas yang diminta oleh manusia bahkan dalam hitungan detik, misalnya *copywriting*, pembuatan teks atau paragraf dengan topik tertentu dan lain sebagainya.

Objek penelitian pada penelitian ini adalah opini masyarakat mengenai aplikasi AI ChatGPT. Data-data untuk objek penelitian akan diambil menggunakan bantuan *library* python yang akan mengumpulkan *tweet* warganet Indonesia di Twitter.

Secara keseluruhan, gambaran penelitian ini dapat dilihat pada Gambar 3.1. Dimulai dari *scraping* *tweet* atau mengumpulkan data dari *tweet*, lalu memberikan label pada data-data opini yang telah dikumpulkan, selanjutnya melakukan *preprocessing* yang meliputi *data cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, berikutnya pembobotan dengan implementasi TF-IDF, selanjutnya dilakukan data split untuk *data train* dan *data test*, lalu dilanjutkan dengan implementasi algoritma *multinomial naive bayes* dan *logistic regression*, selanjutnya data akan di-*train* yang nantinya akan memberikan hasil prediksi dari masing-masing algoritma sehingga dapat dilakukan perbandingan.

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



Gambar 3.1. Gambaran umum penelitian

### 3.2 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah sebagai berikut.

1. **Studi Literatur**, pada tahap ini penulis mengumpulkan literatur dari berbagai pustaka untuk digunakan sebagai penguat serta referensi dari penelitian ini
2. **Analisis Kebutuhan**, pada tahap ini penulis melakukan analisis hal apa saja yang dibutuhkan yang berkaitan dengan penelitian ini, misalnya *library* yang akan digunakan untuk mengumpulkan data yang dibutuhkan
3. **Perancangan Sistem**, Pada tahap ini akan dilakukan perancangan program dan alur *flowchart* yang akan dibuat sesuai dengan analisis kebutuhan dan studi literatur yang telah dilakukan sebelumnya.
4. **Penerapan Sistem**, Pada tahap ini akan dilakukan implementasi dari rancangan yang telah dibuat sebelumnya.
5. **Pengujian dan Evaluasi**, Berikutnya akan dilakukan pengujian untuk memastikan penerapan yang dilakukan berjalan dengan semestinya serta dilakukan evaluasi menggunakan *confusion matrix* untuk mendapatkan hasil performa akurasi, presisi, *recall*, *f1-score* dari dua algoritma yang digunakan serta membandingkan kedua hasilnya.
6. **Penulisan Laporan**, Setelah semua rangkaian metodologi tersebut dilakukan, berikutnya akan dilakukan penulisan karya ilmiah berupa laporan karya ilmiah yang mencakup dokumentasi hasil penelitian.

### 3.3 Perancangan Sistem

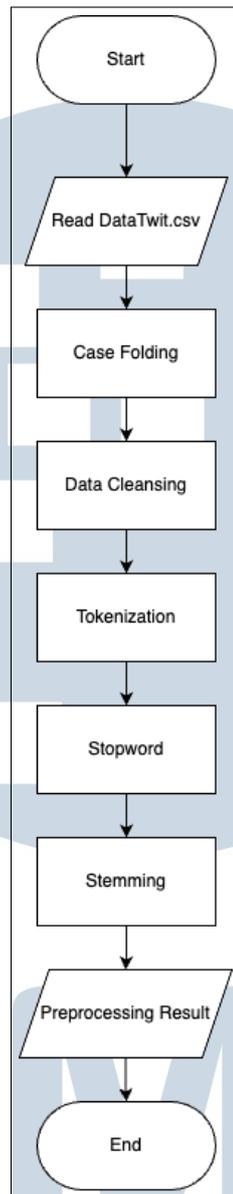
Pada tahap ini akan dirancang proses apa saja yang akan dilakukan serta bagaimana alur dari sistem yang akan diimplementasikan untuk penelitian dalam mengklasifikasi opini terkait ChatGPT dalam sebuah *tweet*. Adapun tahapan yang akan dilakukan untuk melakukan klasifikasi adalah sebagai berikut.

1. *Tweet Crawling* dilakukan untuk mengumpulkan data *tweet* atau *cuitan* dari pengguna Twitter di Indonesia. Pengumpulan *tweet* ini dilakukan dengan menggunakan bantuan *library* *Snsrape* yang menggunakan bahasa pemrograman Python, *library* ini tidak membutuhkan API (Application Programming Interface) resmi dari twitter serta dapat mengumpulkan data

twit yang tidak terbatas. Snsrape juga mendukung pengumpulan data twit pada rentang waktu tertentu seperti halnya penelitian ini yang hanya akan mengambil data twit yang diunggah pada tanggal 30 November 2022 hingga 28 Februari 2023.

2. *Labelling Data* Setelah pengumpulan data dilakukan, berikutnya akan dilakukan labelling yaitu memberikan label pada setiap data yang masuk dengan tujuan mengelompokkan twit tersebut termasuk kelas positif atau negatif. Label positif diberikan kepada twit yang berisi masukan, saran, rasa senang, rasa bahagia, dan emosi positif lainnya, sedangkan kelas negatif berisi kritik, keluhan, rasa marah, rasa tidak suka, serta emosi negatif lainnya. Pada penelitian ini, labelling dilakukan secara otomatis dengan bantuan *library* daftar kata opini berbahasa Indonesia [14].
3. *Text Preprocessing* *Text Proprocessing* merupakan proses yang dilakukan untuk mengolah kata yang sebelumnya tidak terstruktur menjadi terstruktur. Tahapan-tahapan dari proses *text processing* dapat dilihat pada Gambar 3.2.



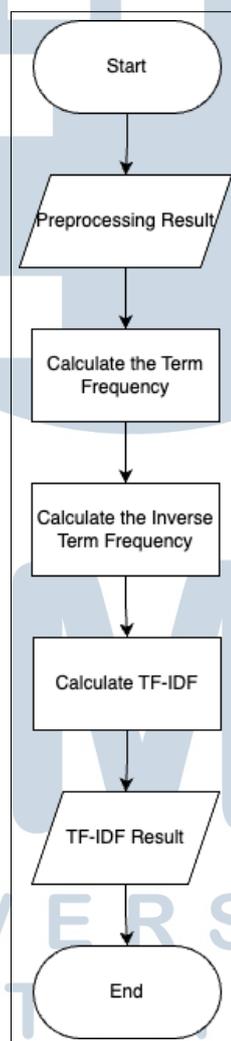


Gambar 3.2. Proses *Text Preprocessing*

Pada Gambar 3.2 tahapan awal yang dilakukan saat melakukan *text preprocessing* adalah membaca file data twit yang berformat CSV, selanjutnya akan dilakukan *case folding* yaitu proses untuk mengubah seluruh huruf kapital pada kalimat menjadi huruf kecil, selanjutnya dilakukan *data cleansing* untuk menghapus angka, simbol, emoji, dan karakter spesial pada kalimat, selanjutnya kata-kata pada kalimat yang sudah melewati *case folding* dan *cleansing* akan disimpan pada array yang disebut token, tahap ini disebut *Tokenization*, selanjutnya akan dilakukan tahap *stopword removal*

yaitu proses penghapusan kata-kata yang tidak memiliki arti dalam sebuah kalimat, tahap terakhir adalah *Stemming* yaitu proses mengubah kata pada teks menjadi kata dasarnya yang sesuai dengan KBBI, tahap *stemming* ini akan menggunakan bantuan library Sastrawi.

4. Implementasi TF-IDF, Implementasi TF-IDF dilakukan untuk menghitung bobot setiap kata yang bermunculan pada dokumen hasil *text-preprocessing* yang telah dilakukan pada tahap sebelumnya. Alur dari penerapan TF-IDF dapat dilihat pada Gambar 3.3 berikut.

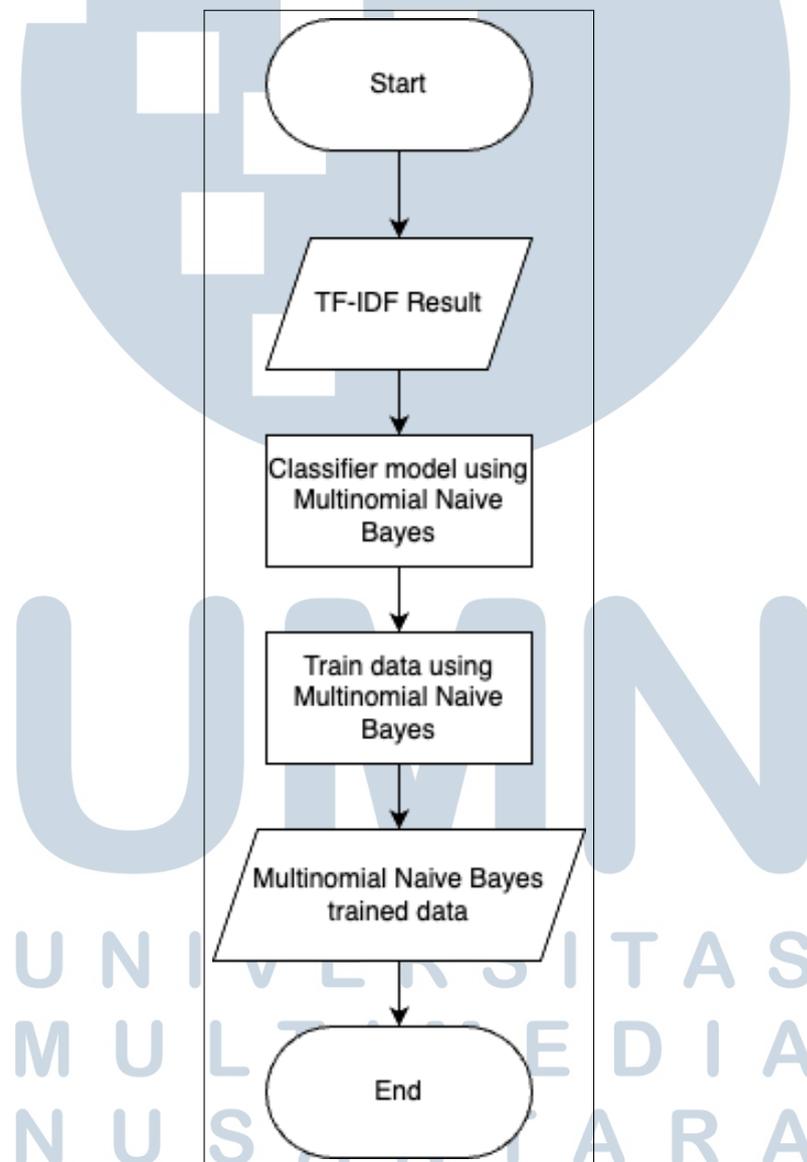


Gambar 3.3. Proses penerapan *TF-IDF*

Sebelum melakukan TF-IDF, data harus sudah melewati tahap *text preprocessing*, selanjutnya dilakukan penghitungan *Term Frequency* terlebih dahulu, selanjutnya dilakukan penghitungan *Inverse Term Frequency*.

Setelah penghitungan nilai TF dan IDF dilakukan, selanjutnya dilakukan penghitungan TF-IDF yang nantinya akan menghasilkan nilai TF-IDF.

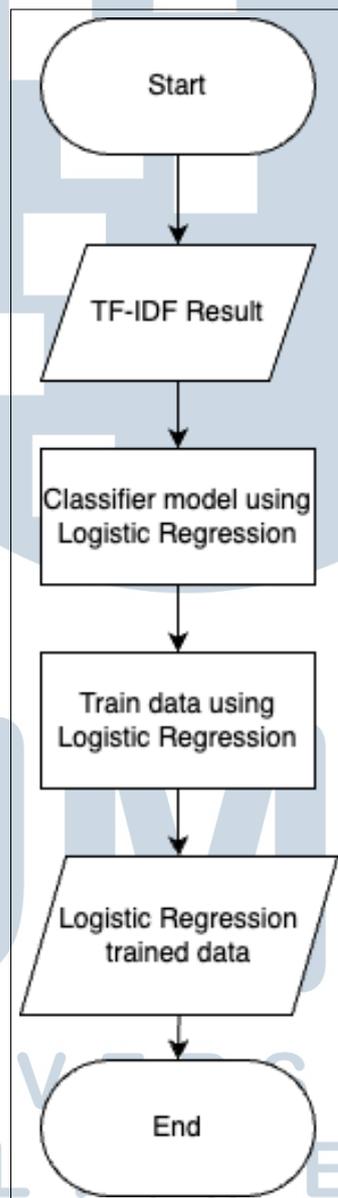
5. Implementasi Multinomial Naive Bayes, Setelah tahap implementasi TF-IDF dilakukan, selanjutnya dilakukan implementasi Multinomial Naive Bayes yang menggunakan data hasil implementasi TF-ID yang telah dibagi dua bagian, yaitu *data train* dan *data test*, selanjutnya *data train* akan di-*train* dan *data test* akan diprediksi.



Gambar 3.4. Proses penerapan *multinomial naive bayes*

6. Implementasi Logistic Regression, Tahapan ini serupa dengan tahapan

sebelumnya hanya saja pada tahapan ini algoritma yang digunakan untuk train dan test adalah *logistic regression*. Setelah data nilai *matrix confusion* dari klasifikasi Multinomial Naive Bayes dan Logistic Regression didapatkan, selanjutnya akan dilakukan perbandingan akurasi, presisi, *recall*, dan *f1-score* dari masing-masing algoritma.



Gambar 3.5. Proses penerapan *logistic regression*