BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian ini adalah analisis sentimen publik terhadap gerakan boikot Israel yang dilakukan di *platform X* (sebelumnya Twitter). Sentimen publik yang dianalisis mencakup opini, pandangan, serta reaksi pengguna terhadap gerakan boikot Israel, yang diekspresikan melalui postingan atau komentar di platform tersebut.

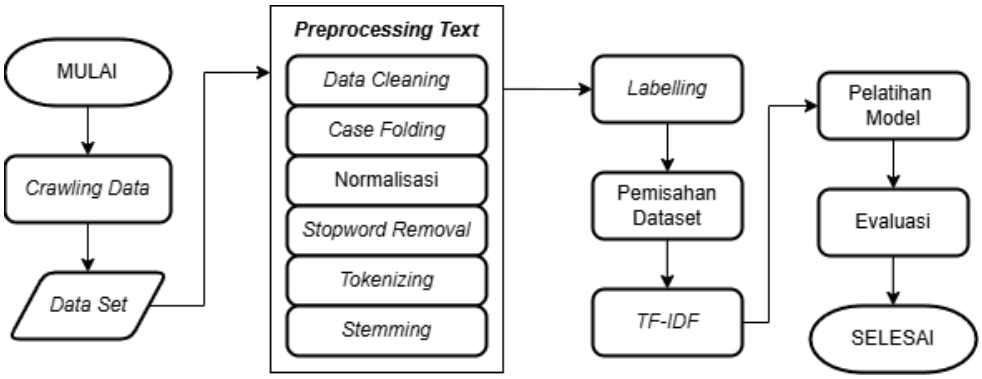
Data yang digunakan dalam penelitian ini berupa tweet publik yang mengandung kata kunci terkait gerakan boikot Israel. Data ini akan dikumpulkan dalam periode tertentu untuk memperoleh gambaran yang lebih representatif mengenai sentimen masyarakat.

Penelitian ini menggunakan metode *machine learning*, dengan algoritma *Random Forest* dan *Logistic Regression* sebagai model klasifikasi utama untuk mengelompokkan sentimen menjadi positif, negatif, dan netral. Penggunaan kedua algoritma ini bertujuan untuk membandingkan performa masing-masing model dalam menganalisis sentimen publik.

Dengan demikian, penelitian ini berfokus pada penerapan dan evaluasi algoritma *Random Forest* dan *Logistic Regression* dalam memahami pola sentimen publik terkait gerakan boikot Israel di platform X.

3.2 Prosedur Penelitian

Pada prosedur penelitian menjelaskan tahapan-tahapan yang dilakukan dalam penelitian untuk mencapai tujuan yang telah ditetapkan. Prosedur penelitian dirancang secara sistematis agar setiap langkah dapat dilakukan dengan runtut dan terstruktur. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap gerakan boikot israel di *platform X* menggunakan algoritma *Random Forest* dan *Logistic Regression*. Oleh karena itu, proses penelitian mencakup serangkaian langkah mulai dari pengumpulan data, pengolahan data, hingga analisis hasil prediksi sentimen. Berikut merupakan gambar dari alur penelitian:



Gambar 3. 1 Alur Penelitian

3.2.1 *Crawling Data*

Proses pengumpulan data dilakukan dari *platform X* menggunakan Teknik *Web Crawling* dengan memanfaatkan *auth token* dari akun *X*, Proses *crawling* dilakukan untuk mengumpulkan *tweet-tweet* yang berisi opini atau sentimen terkain gerakan boikot Israel. Kata kunci seperti Boikot, Israel, Palestine digunakan sebagai parameter pencarian untuk menghasilkan data yang sesuai dengan topik penelitian.

Proses *crawling* dimulai dari autentikasi menggunakan *auth token*, yang memungkinkan akses ke konten publik pada *platform X*. data yang dikumpulkan meliputi teks *tweet*, tanggal, dan waktu posting, serta informasi tambahan seperti jumlah *likes* dan *retweet* jika diperlukan. Rentang waktu pengumpulan data dibatasi pada periode tertentu yaitu dari tanggal 1 Maret 2023 sampai 30 Januari 2025 hal itu untuk memastikan relevansi data dengan isu yang sedang dibahas.

Data hasil crawling memperoleh 616 data lalu disimpan dalam bentuk format CSV, yang memudahkan analisis lebih lanjut. Dalam hal ini, perhatian khusus diberikan pada etika penelitian, termasuk menjaga kerahasiaan data pengguna dan menjamin bahwa data yang diperoleh dipakai hanya untuk kepentingan penelitian. Tahap preprocessing dilakukan menggunakan data dari proses crawling terebut dan digunakan untuk menganalisis sentimen untuk mengidentifikasi opini publik terhadap gerakan boikot Israel menggunakan model *Random Forest* dan *Logistic Regression.*

3.2.2 *Preprocessing Data*

Proses *preprocessing* meliputi *data cleaning, case folding,* normalisasi, *stopword removal*, tokenisasi, dan *stemming*. *Data cleaning* membersihkan karakter yang tidak diperlukan, *case folding* mengganti semua buruf kapital menjadi huruf kecil, dan normalisasi mengganti kata tidak baku menjadi baku. *Stopword removal* menghilangkan kata-kata umum yang kurang relevan, tokenisasi memecah kalimat menjadi kata-kata, dan *stemming* mengubah kata ke bentuk dasarnya. Setelah dilakukan *preprocessing data*, jumlah data yang tersisa yaitu 616 data. Tahapan ini bertujuan untuk menyederhanakan dan menstandarisasi data agar analisis teks lebih akurat.

3.2.3 *Labelling*

Proses pelabelan data dalam penelitian ini dilakukan secara manual dengan bantuan seorang ahli pakar bahasa Indonesia, yaitu Adelya Dariyah, M.Pd. Beliau berperan penting dalam mengklasifikasikan sentimen pada data tweet menjadi tiga kategori, yaitu sentimen positif, negatif, dan netral. Penilaian dilakukan dengan mempertimbangkan konteks bahasa, makna tersirat, serta penggunaan diksi dalam setiap tweet yang telah dikumpulkan. Keterlibatan ahli bahasa ini bertujuan untuk meningkatkan validitas dan akurasi pelabelan data, sehingga model yang dibangun dapat melakukan prediksi sentimen dengan lebih baik dan sesuai dengan karakteristik bahasa Indonesia yang digunakan di media sosial.

3.2.4 *Pemisahan Dataset*

Proses pemisahan data dibagi menjadi 2 yaitu data latih dan data uji yang bertujuan untuk membagi data dari hasil *preprocessing* ke dalam perbandingan tertentu. Jumlah data yang diperoleh dari preprocessing ada 616 lalu dibagi dengan perbandingan jumlah pembagian data dalam penelitian ini yaitu 80%:20% terdiri dari 80% data train sebanyak 492 dan 20% data test sebanyak 124. Pembagian dengan perbandingan tersebut dikarenakan semakin banyak data yang dilatih, maka akan memperoleh hasil analisis lebih akurat.

3.2.5 *Pembobotan TF-IDF*

Metode pembobotan kata dalam dokumen atau sering disebut juga TF-IDF bertujuan untuk mengonversi teks ke format numerik dan meningkatkan akurasi analisis. Perhitungan TF-IDF melibatkan Term Frequency atau TF, yang mengukur

frekuensi kata yang muncul pada dokumen, serta Inverse Document Frequency atau IDF, yang memberi bobot lebih kecil pada kata umum di seluruh dokumen. Nilai IDF dihitung menggunakan logaritma berdasarkan jumlah total dokumen dan dokumen yang mengandung kata tertentu. Pelatihan Model

Penelitian ini menggunakan model Random Forest dan Logistic Regression guna menganalisis sentimen publik terhadap gerakan boikot Israel di platform X. Random Forest membangun beberapa pohon keputusan untuk menghasilkan prediksi yang lebih akurat, sementara Logistic Regression memodelkan hubungan antara fitur dan label sentimen menggunakan fungsi logit. Fitur teks dikonversi ke bentuk numerik, seperti TF-IDF, sebelum digunakan sebagai input model. Model dilatih pada data latih dan di evaluasi dengan data uji menggunakan metrik akurasi, precision, recall, f1-score untuk memastikan kinerja prediksi yang optimal.

3.2.6 *Confusion Matrix*

Confusion matrix yaitu metode evaluasi model klasifikasi untuk membandingkan hasil prediksi dengan label asli dalam bentuk tabel. Terdapat empat elemen utama: True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). Dari matriks ini, dapat dihitung metrik evaluasi seperti accuracy, precision, recall, dan f1 score.

Pada riset ini, proses confusion matrix bertujuan untuk evaluasi model Random Forest dan Logistic Regression dalam analisis sentiment gerakan boikot israel di platform X.

Dengan matriks ini, dapat diketahui letak kesalahan prediksi serta performa keseluruhan model. Berikut merupakan rumus dari confusion matrix :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (4)$$