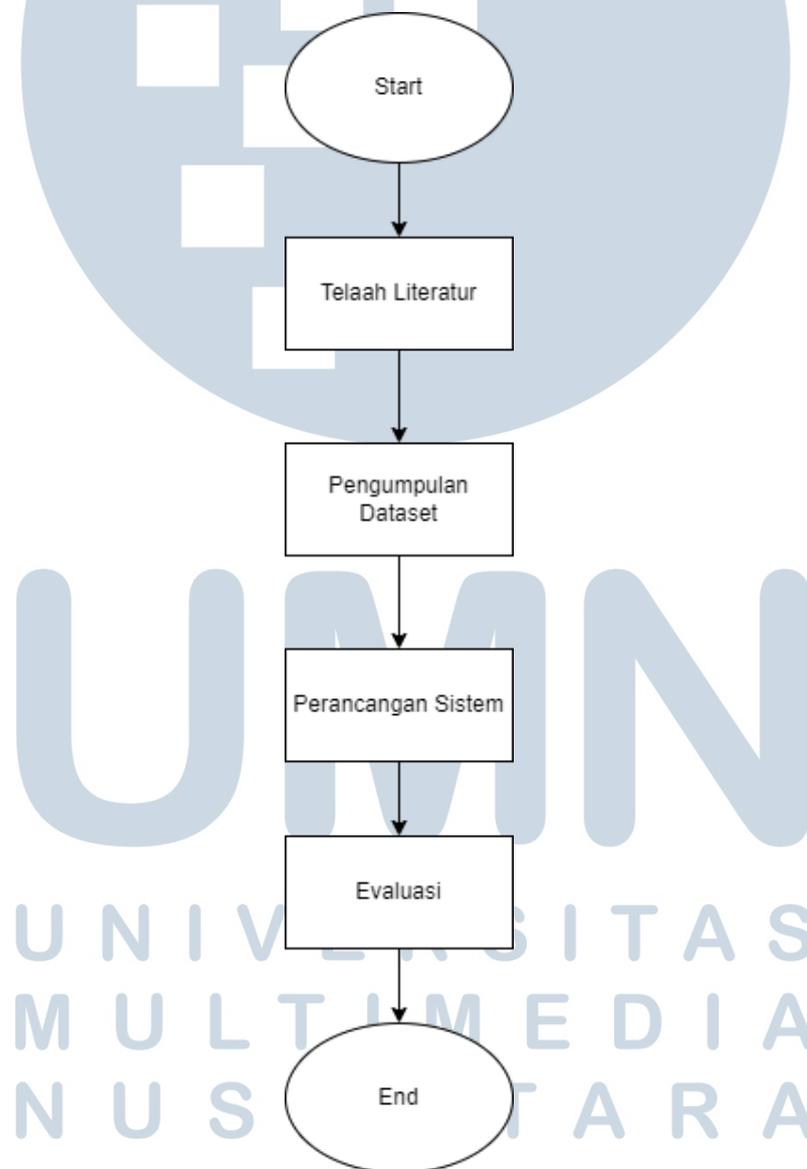


BAB 3 METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Pengerjaan penelitian "Implementasi Algoritma Decision Tree untuk Penyaringan Komentar Negatif di Instagram dilaksanakan secara bertahap. Tahap-tahap pengerjaan penelitian ini dijabarkan seperti Gambar 3.1 dibawah ini.



Gambar 3.1. Metodologi Penelitian

3.1.1 Telaah Literatur

Pada tahapan ini, dilakukan riset dan pembelajaran dari beberapa referensi dan jurnal yang akan digunakan untuk dalam penelitian ini. Teori yang di pelajari seperti *decision tree*, *CART algorithm* dan sebagainya.

3.1.2 Pengumpulan Dataset

Pada tahapan ini, dilakukan pengumpulan data yang akan dipakai untuk melakukan penelitian ini. Data yang dipakai adalah kumpulan contoh komentar user di *instagram*. Data tersebut akan diambil dari *github*. Dataset yang dipakai adalah *dataset hate speech insatgram* dari *github* yang berisi 572 contoh komentar di *instagram*.

3.1.3 Perancangan Sistem

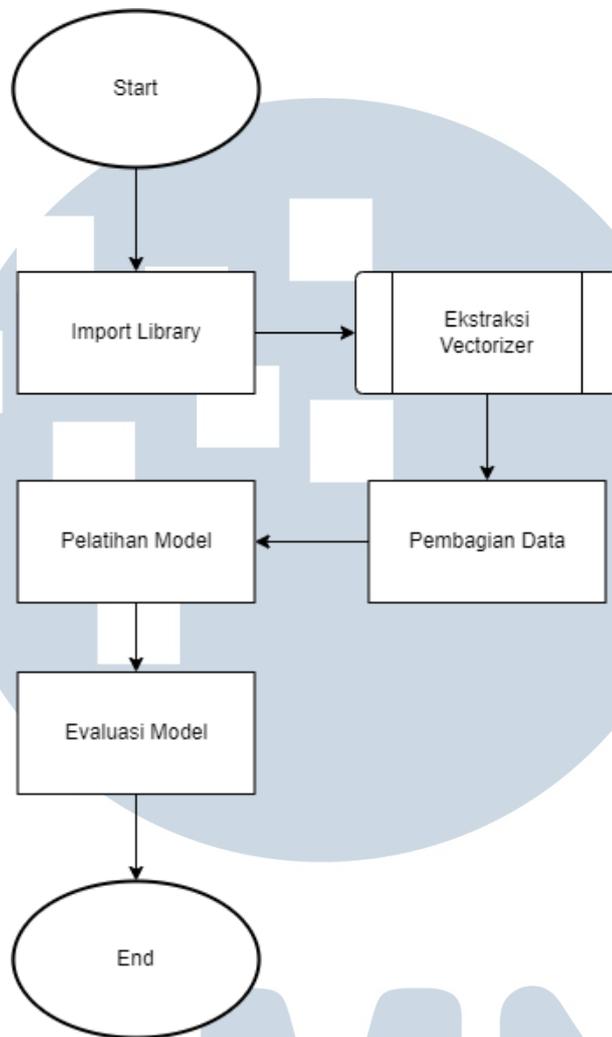
Pada tahapan ini, dilakukan perancangan sistem dan membuat model untuk mengolah dataset yang sudah dikumpulkan dengan menggunakan algoritma *decision tree*

3.1.4 Evaluasi

Setelah model berhasil dibuat, dilakukan pengujian dimana model digunakan untuk melakukan klasifikasi komentar dari dateset yang telah disiapkan. Hasil pengujian tersebut dievaluasi mengenai performa, akurasi, dan hasilnya oleh sistem *Confusion Matrix*.

3.2 Processing Data

Tahapan *Processing Datayang* ada dalam penelitian diuraikan pada bagian ini. Untuk menggambarkan alur *Processing Data*, digunakan *flowchart* yang ada pada Gambar 3.2.



Gambar 3.2. Flowchart Processing Data

3.2.1 Import Library

Pada tahapan ini, proses dimulai dengan mengimpor beberapa modul penting dari *library sklearn*, termasuk *train test split* untuk membagi data menjadi set pelatihan dan pengujian, *TfidfVectorizer* untuk mengubah teks menjadi format numerik menggunakan metode *TF-IDF*, *DecisionTreeClassifier* sebagai algoritma klasifikasi, dan *accuracy score* untuk mengukur akurasi prediksi model. Data yang digunakan terdiri dari komentar dan label yang sesuai, di mana komentar adalah teks yang akan diklasifikasikan dan label menunjukkan kategori dari teks tersebut, misalnya 1 untuk positif dan 0 untuk negatif. Sebuah pernyataan *assert* digunakan

untuk memastikan bahwa jumlah komentar dan label harus sama, menghindari kesalahan dalam pemrosesan data.

3.2.2 Ekstraksi Vectorizer

Pada tahapan ini, *TfidfVectorizer* digunakan untuk mengubah setiap komentar menjadi vektor TF-IDF. Proses ini mengubah teks menjadi format yang dapat dipahami dan diproses oleh model pembelajaran mesin, dengan setiap kata diwakili sebagai dimensi dalam ruang fitur dan bobotnya menunjukkan kepentingannya.

3.2.3 Pembagian Data

Pada tahapan ini, Data kemudian dibagi menggunakan *train test split*, di mana 33 persen dari data dijadikan sebagai set pengujian dan sisanya sebagai set pelatihan, dengan *randomstate* yang ditetapkan untuk memastikan konsistensi pembagian data.

3.2.4 Pelatihan Model

Pada tahapan ini, Model *DecisionTreeClassifier* dilatih dengan set pelatihan. Dalam pembelajaran, model mencoba memahami pola dan hubungan antara fitur dan label. *Decision Tree* bekerja dengan membuat struktur pohon di mana setiap *node* mewakili keputusan yang dibuat berdasarkan nilai fitur. Setelah model dilatih, ia digunakan untuk memprediksi label pada set pengujian. Akurasi dari prediksi ini dihitung dengan membandingkan prediksi dengan label sebenarnya, dan hasilnya dicetak ke konsol. Akurasi ini memberikan gambaran tentang seberapa efektif model dalam mengklasifikasikan komentar berdasarkan konten teksnya.

3.2.5 Evaluasi Model

Akhir dari tahapan *processing data* pada penelitian ini adalah mengevaluasi dan menganalisa kinerja kedua model yang telah dibangun. Proses ini bertujuan untuk memperoleh informasi dan pemahaman lebih terhadap kelebihan dan kekurangan dari penerapan model *Decision Tree*.