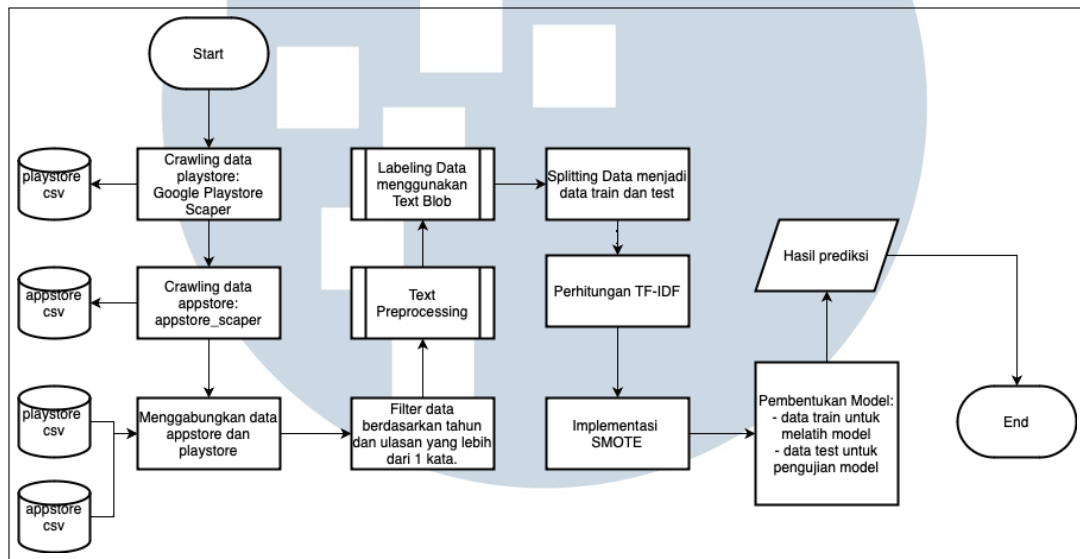


## BAB 3 METODOLOGI PENELITIAN

### 3.1 Alur Penelitian

Penelitian ini mengikuti langkah-langkah yang diuraikan dalam bentuk *flowchart* pada Gambar 3.1.



Gambar 3.1. *Flowchart* alur penelitian.

### 3.2 Metodologi Penelitian

Pada bagian ini akan membahas tentang metode yang akan dilakukan selama penelitian berlangsung dimana dimulai dengan studi literatur, perancangan sistem, pengujian model, dan evaluasi dai model yang telah dibentuk.

#### 3.2.1 Studi Literatur

Pada tahapan ini mencari berbagai sumber informasi dari jurnal penelitian yang relevan dengan studi yang akan dilakukan. Analisis kebutuhan dilakukan pada tahap ini untuk menganalisis kebutuhan akan data, metode, dan prosedur yang diperlukan dalam penelitian ini, seperti cara pengumpulan data, tahapan *text preprocessing*, cara perhitungan TF-IDF, penggunaan SMOTE, dan cara kerja algoritma *Naïve Bayes*.

### 3.2.2 Perancangan Sistem

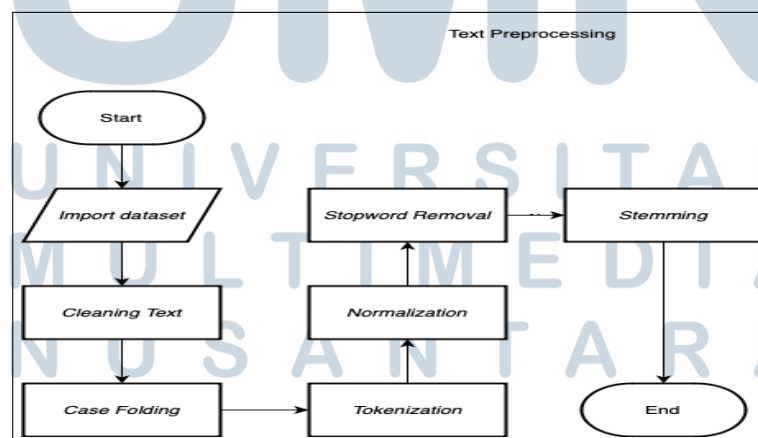
Pada tahap ini akan melakukan perancangan sistem dimulai dengan melakukan *crawling data* dari *appstore* dan *play store*, kemudian dilakukan *text-preprocessing* yang meliputi tahapan *cleaning text*, *case folding*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*. Kemudian, dilakukan *labeling* data menggunakan *library TextBlob*. Setelah itu, model akan dibentuk menggunakan *Naïve Bayes* dan model tersebut akan diuji dan dievaluasi.

#### A Data Collection

Pada tahapan ini akan melakukan pengumpulan data terkait ulasan pengguna terhadap kualitas pelayanan yang diberikan dalam aplikasi Alfigift melalui dua *platform*, yaitu *play store* dan *app store*. Pengumpulan data dari *play store* akan menggunakan *Google Play Scraper*, sedangkan pengumpulan data ulasan pada *app store* akan menggunakan *app.store.scraper*. Periode data diambil dari tanggal 1 Januari 2023 sampai 1 Mei 2024, karena untuk dapat mengamati perubahan tren, sikap, dan emosi pengguna terhadap aplikasi yang sangat penting untuk pengembangan aplikasi kedepannya.

#### B Text Pre-processing

Pada proses ini akan dilakukan pemilihan dan pembersihan teks. Hal ini bertujuan untuk membuat teks menjadi lebih terstruktur dan efisien. Proses langkah-langkah *text preprocessing* akan mengikuti pada Gambar 3.2.



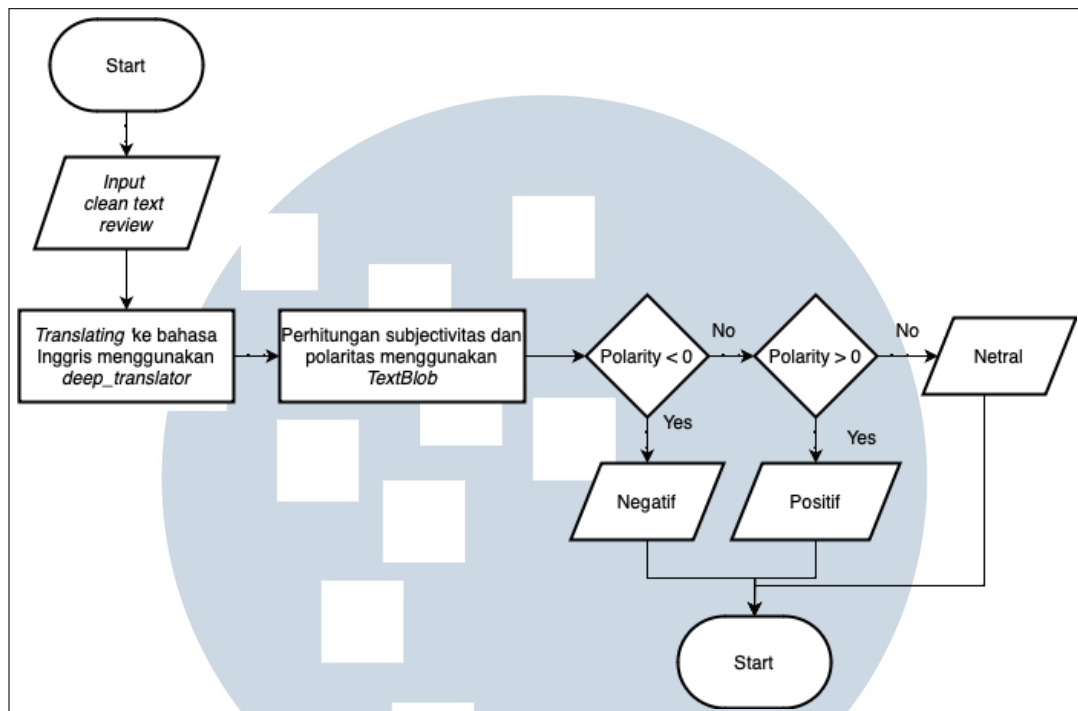
Gambar 3.2. Flowchart text preprocessing.

1. *Cleaning Text*, merupakan tahap dimana data teks dibersihkan dari simbol-simbol dan tanda baca yang tidak diperlukan.
2. *Case Folding*, merupakan langkah untuk mengubah semua huruf dalam kalimat menjadi huruf kecil.
3. *Tokenization*, merupakan proses membagi sekumpulan karakter dalam teks menjadi unit kata.
4. *Normalization*, merupakan langkah untuk mengubah kata *slang* menjadi kata baku.
5. *Stopword*, merupakan langkah untuk menghilangkan kata-kata umum yang sering muncul dan tidak memiliki makna.
6. *Stemming*, merupakan proses mereduksi kata-kata menjadi bentuk dasarnya.

### **C Labeling Data**

Proses pelabelan data dilakukan secara otomatis dengan menggunakan *library TextBlob*. Proses ini akan dilakukan dengan mengambil ulasan dari pengguna Alfabeta, dimana ulasan tersebut yang rata-rata menggunakan bahasa Indonesia akan dikonversikan menjadi bahasa Inggris, karena *library TextBlob* hanya menerima data yang berbahasa Inggris. Kemudian, akan dilihat berdasarkan *polarity* dan *subjectivity*, jika hasil perhitungan analisa menunjukkan skor 0 maka ulasan tersebut ditandai sebagai sentimen netral, sedangkan lebih dari 0 maka akan ditandai sebagai sentimen positif, dan jika kurang dari 0 maka akan ditandai sebagai sentimen negatif. Tahapan *labeling* data ditunjukkan pada Gambar 3.3.

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 3.3. Flowchart labeling data.

#### D Data Splitting

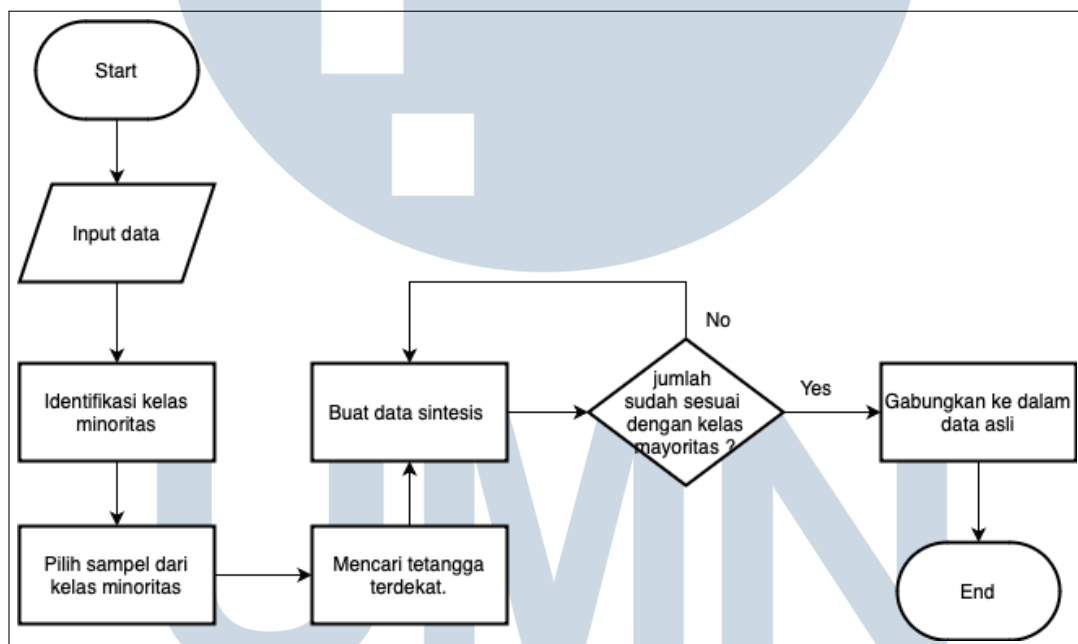
Data akan dipisahkan menjadi 8:2, dimana 80% data akan digunakan sebagai data *training* dalam pembentukan model dan 20% akan digunakan sebagai data *testing* untuk model yang telah dibentuk. Proses pembagian data ini dilakukan dengan, 80% data dari masing-masing kelas akan diambil sebagai data *training* dan 20% dari masing-masing kelas akan diambil sebagai data *testing*.

#### E Implementasi TF-IDF

Implementasi TF-IDF digunakan untuk mengubah data teks yang telah didapatkan menjadi bentuk numerik, sehingga data teks tersebut dapat diolah dan diproses oleh mesin. Penggunaan TF-IDF juga dapat membantu untuk memprioritaskan kata-kata yang penting dan signifikan dalam memberikan konteks positif, negatif, dan netral.

## F Implementasi SMOTE

Implementasi SMOTE digunakan untuk menyeimbangkan data antar label positif, negatif, dan netral. Hal ini dilakukan agar pembelajaran model tidak menjadi bias terhadap kelas yang minoritas. Implementasi SMOTE juga dapat meningkatkan kinerja model dalam mendeteksi sentimen dari kelas yang minoritas. Pada Gambar 3.4 menunjukkan cara kerja SMOTE yang dimulai dengan identifikasi data minoritas dalam *dataset*, kemudian SMOTE akan mencari tetangga yang terdekat dan akan dibentuk data sintetik dengan memilih salah satu tetangga terdekat dan menambahkan titik baru diantara sampel asli dan tetangganya. Hal ini dilakukan berulang sampai data kelas minoritas mencapai data kelas mayoritasnya.

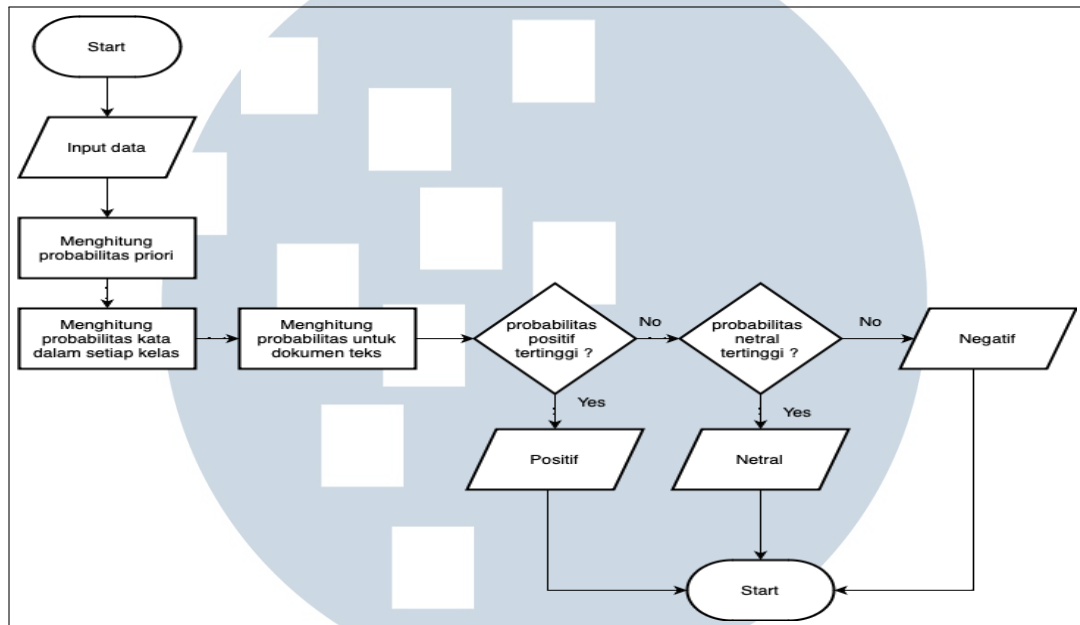


Gambar 3.4. Flowchart implementasi smote.

## G Pembentukan Model menggunakan Algoritma Naïve Bayes

Data yang telah dilakukan diproses akan dianalisa dan diklasifikasikan menggunakan algoritma *Multinomial Naïve Bayes*. Model akan dihitung menggunakan rumus *Multinomial Naïve Bayes*, dimana jika data memiliki probabilitas kelas terhadap sentimen positif lebih akan dimasukkan ke dalam kelompok positif, sedangkan jika probabilitas kelas negatif lebih tinggi akan

dimasukkan ke dalam kelompok negatif, dan jika probabilitas kelas netral lebih tinggi akan masuk ke dalam kelompok netral. Langkah-langkah pembentukkan model mengikuti *flowchart* pada Gambar 3.5.



Gambar 3.5. *Flowchart* pembentukan model *Naive Bayes*.

### 3.2.3 Pengujian model

Dalam pengujian model terdapat dua skenario, yaitu data akan dibagi dengan rasio 80% data *training* dan 20% data *testing*, kemudian skenario kedua, data akan dibagi dengan rasio 80% data *training* dan 20% data *testing*. Hasil dari pengujian ini akan menghasilkan laporan yang mencakup *presisi*, *recall*, dan *f-1 score* untuk data yang bersentimen positif dan negatif. Selain itu, akurasi total juga dihitung dan disajikan dalam laporan.

### 3.2.4 Evaluasi

Proses evaluasi akan membandingkan antara hasil perhitungan *accuracy*, *precision*, *recall*, dan *f1-score* yang telah dilakukan oleh mesin pada model yang dibentuk dengan perhitungan manual dari *confusion matrix* yang telah didapatkan dari hasil pengujian model.