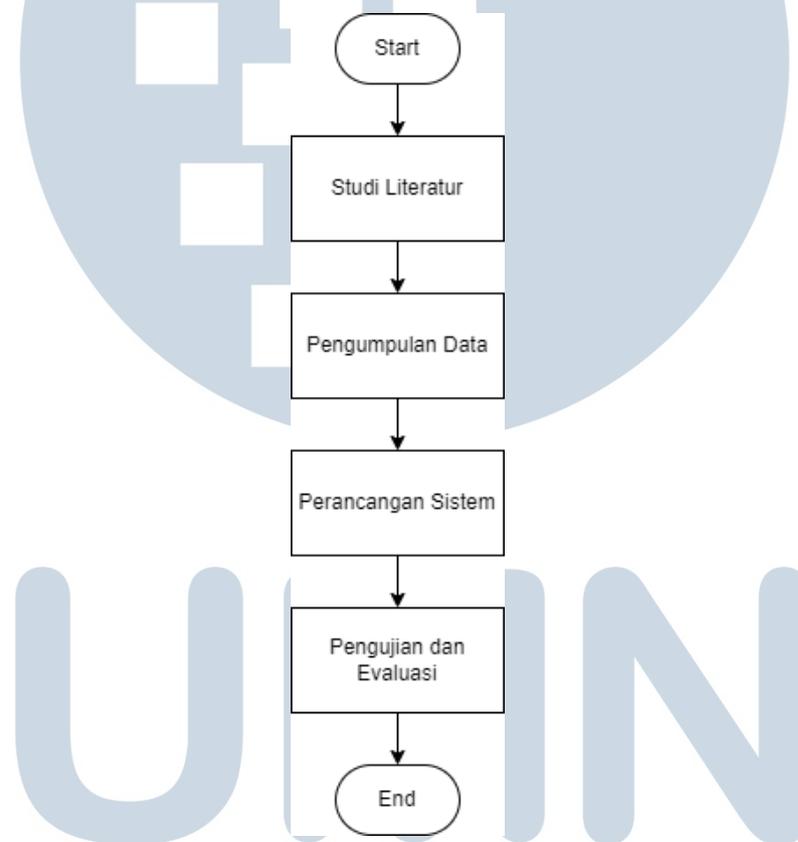


BAB 3 METODOLOGI PENELITIAN

3.1 Alur Metodologi Penelitian

Penelitian yang dilakukan mengacu pada metodologi atau langkah-langkah yang digambarkan dalam sebuah bentuk *flowchart*. Gambar 3.1 merupakan diagram alir atau *flowchart* metodologi penelitian ini.



Gambar 3.1. Diagram alir metodologi penelitian

3.1.1 Studi Literatur

Studi literatur dilakukan guna mempelajari dan memahami literatur dan teori yang digunakan dalam penelitian. Pada penelitian ini, dibutuhkan literatur mengenai Twitter, analisis sentimen, dan metode Naïve Bayes.

3.1.2 Pengumpulan Data

Pengumpulan data merupakan langkah awal dalam praktek penelitian ini. Data diambil menggunakan *library snsrape*. Data yang diambil berisi teks yang berhubungan dengan *keywords* seputar kenaikan harga BBM di Indonesia, dengan lama satu tahun sebelum data dikumpulkan. Pengumpulan data dilakukan dengan menggunakan *library Snsrape* pada hari Minggu, 16 April 2023.

3.1.3 Perancangan Sistem

Perancangan sistem dilakukan setelah seluruh data-data yang dibutuhkan telah terkumpul. Data yang telah *dimining* diolah melalui tahap *text pre-processing* untuk dibersihkan dan dikonsolidasi ke dalam bentuk data yang sesuai.

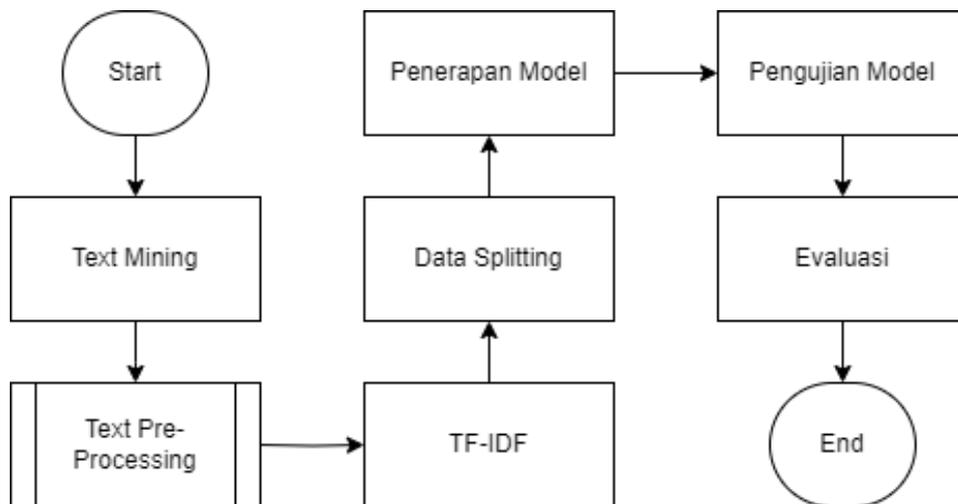
3.1.4 Pengujian dan Evaluasi

Setelah model melalui tahap *training* dengan baik, dilakukan pengujian dimana model digunakan untuk melakukan klasifikasi sentimen dari *test data* yang telah disiapkan. Hasil pengujian tersebut dievaluasi mengenai performa, akurasi, dan hasilnya oleh sistem *Confusion Matrix*.

3.2 Rancangan Sistem

Tahapan rancangan yang ada dalam penelitian diuraikan pada bagian ini. Untuk menggambarkan alur rancangan sistem, digunakan *flowchart* yang ada pada Gambar 3.2.

U M M N
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 3.2. Diagram Alir Rancangan Sistem

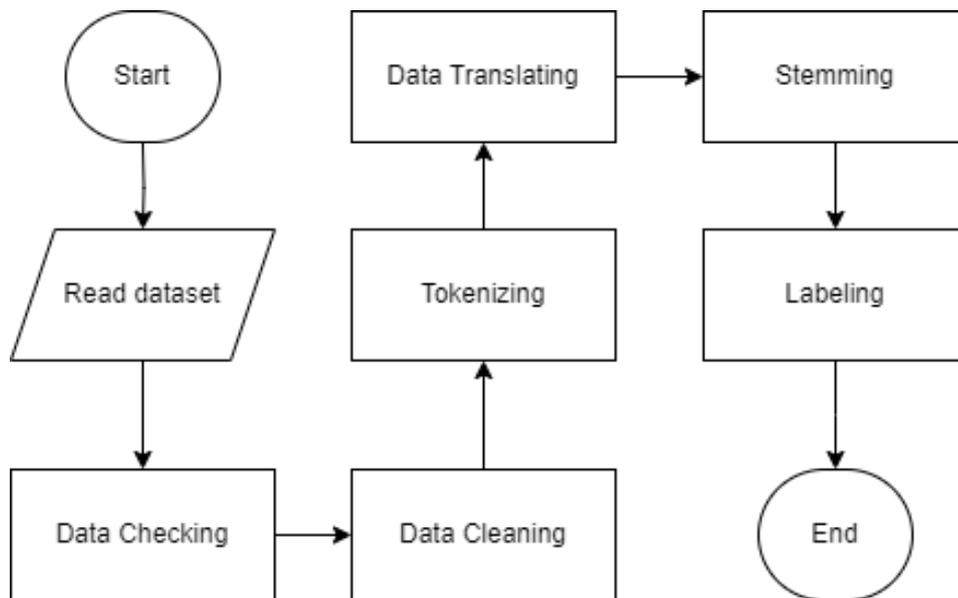
3.2.1 Text Mining

Data yang menjadi bagian besar dalam penelitian analisis sentimen ini diambil menggunakan *library Snsrape*. *Library* ini memungkinkan pengguna untuk mengambil data *tweets* dari pengguna Twitter dengan keyword yang ditentukan. Pada penelitian analisis sentimen ini, data berfokus terhadap kenaikan harga BBM yang terjadi di Indonesia. Oleh karena itu, *keyword* yang menjadi acuan *Text mining* adalah "Kenaikan Harga BBM", karena *keyword* berbahasa Indonesia memperbesar banyaknya data berbahasa dan berlokasi di Indonesia.

3.2.2 Text Pre-Processing

Data kemudian diolah dan dibersihkan agar model dapat terlatih dengan data bersih. Tahap ini mencakup beberapa langkah-langkah yang digambarkan dalam bentuk *flowchart* pada Gambar 3.3.

UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3.3. Diagram Alir *Text Pre-Processing*

Proses diawali dengan mengimpor dan membaca dataset yang telah didapatkan melalui proses *text mining*. Seluruh karakter huruf pada data dikonversi menjadi huruf kecil. Kemudian data diperiksa jika ada data yang *NoneType* atau terduplikat. Jika ada data yang *NoneType*, maka data *didrop*. Data *NoneType* dihapus karena bukan merupakan karakter yang mengandung sentimen dan bisa mengakibatkan *error* pada proses *data splitting*. *Error* tersebut dapat terjadi karena data *NoneType* tidak memiliki arti, sehingga tidak dapat dilakukan *splitting* antara data yang mengandung karakter dengan data yang *NoneType*. Data yang terduplikat juga dihapus duplikasinya sehingga seluruh data unik dan berbeda satu sama lain. Penghapusan duplikasi diperlukan karena duplikasi dapat menimbulkan kekeliruan terhadap hasil dari klasifikasi yang dilakukan oleh model. Hasil dari model yang dilatih menggunakan data *train* yang memiliki duplikasi bisa mendapatkan hasil yang berbeda ketika melakukan prediksi terhadap data *test* yang tidak memiliki duplikasi.

Data kemudian dibersihkan dari angka, simbol, dan tanda baca sehingga menyisakan kata dengan *character*. Data yang telah bersih *ditokenize* untuk memecah kata menjadi bentuk token. Data kemudian diperiksa bahasanya, kemudian jika ada bahasa selain bahasa Indonesia, diterjemahkan dengan menggunakan *library googletrans*. Setelah melalui proses *translating*, data melalui proses *stemming* yang mengubah kata-kata dalam data menjadi kata dasar. Dengan demikian, proses selanjutnya yaitu *labelling*, menjadi lebih akurat karena

menggunakan kata baku yang memiliki kemungkinan lebih besar untuk berada pada kamus yang digunakan sebagai acuan skor label.

Proses *labelling* merupakan proses pemberian label untuk setiap baris data. Setiap data memiliki kata-kata yang dapat dikalkulasikan nilai skor sentimennya dengan mencocokkan data yang telah *distemming* terhadap kamus khusus untuk *labelling*. Masing-masing skor pada tiap data dijumlahkan dan diberikan label sesuai dengan ketentuan:

- Skor data lebih dari 0, sentimen dianggap positif
- Skor data kurang dari 0, sentimen dianggap negatif

3.2.3 TF-IDF

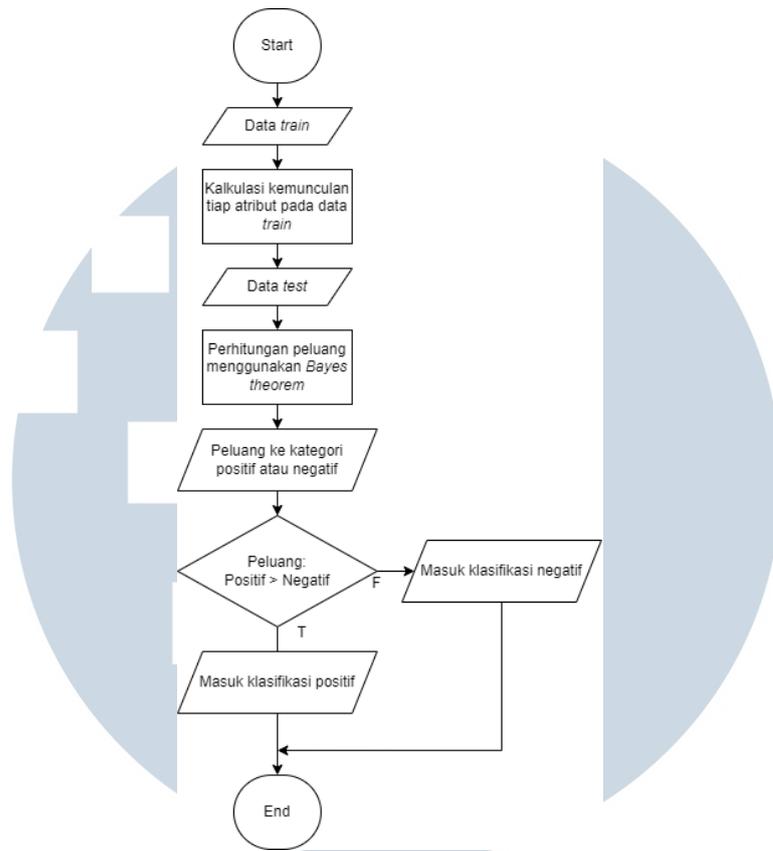
Setelah data memiliki label masing masing, dilakukan pembobotan *TF-IDF*. Proses ini merupakan pembobotan statistik terhadap data untuk mengukur seberapa penting suatu kata pada dokumen data tersebut.

3.2.4 Data Splitting

Data dipisahkan antara data *train* dan data *test*. Model dilatih menggunakan data *train*, kemudian diuji dengan data *test*. Sebanyak 20% data *test* menjadi bahan uji terhadap model *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* yang digunakan pada penelitian ini.

3.2.5 Penerapan Model Multinomial Naïve Bayes dan Complement Naïve Bayes

Penelitian menggunakan dua jenis model Naïve Bayes, yaitu *Multinomial Naïve Bayes* dan *Complement Naïve Bayes*. Kedua model dilatih menggunakan data *train* yang sama dengan skenario *real data* bersamaan dengan pembobotan TF-IDF. Model diimplementasikan menggunakan bahasa pemrograman *Python 3* dengan cara kerja yang digambarkan dalam bentuk diagram alir pada Gambar 3.4.



Gambar 3.4. Diagram Alir Cara Kerja Model *Naïve Bayes*

Model bekerja dengan cara melakukan kalkulasi terhadap jumlah kemunculan atribut pada kategori tertentu [38]. Perhitungan dilakukan terhadap data *train* sebagai acuan utama untuk melakukan klasifikasi terhadap data yang baru. Pada penelitian ini, model diuji menggunakan data *test*, sehingga dilakukan uji klasifikasi terhadap data *test*. Model menggunakan rumus *Bayes theorem* untuk menghitung peluang masuknya data ke kategori positif atau negatif. Jika data memiliki kecenderungan positif, maka data diklasifikasikan sebagai positif. Jika data memiliki kecenderungan negatif, maka data diklasifikasikan sebagai negatif.

3.2.6 Pengujian Model

Setelah model terlatih, dilakukan pengujian terhadap kedua model dengan data *test* yang sama menggunakan *confusion matrix*. Hasil dari pengujian tersebut berupa *report* mengenai *precision*, *recall*, dan *f-1 score* untuk data bersentimen positif dan negatif. Total akurasi juga dihitung dan dipaparkan dalam *report*.

3.2.7 Evaluasi

Akhir dari tahapan rancangan sistem 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 *Naïve Bayes*. Informasi ini diharapkan berguna untuk para pembaca.

