

BAB 3

METODOLOGI DAN PERANCANGAN SISTEM

3.1 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah sebagai berikut:

1. **Pemahaman Literatur**, pada tahap ini mengumpulkan berbagai sumber informasi dari jurnal penelitian yang berkaitan dengan penelitian yang akan dilakukan.
2. **Analisa Kebutuhan**, pada tahap ini melakukan analisis kebutuhan terhadap data, metode, dan prosedur yang diperlukan pada penelitian ini.
3. **Desain Sistem**, pada tahap ini membuat rancangan alur program menggunakan *flowchart* dari hasil pengumpulan literatur dan analisis kebutuhan yang sudah dilakukan.
4. **Pemrograman Sistem**, pada tahap ini merealisasikan proses implementasi dari rancangan yang sudah dibuat pada tahapan sebelumnya ke dalam bentuk *coding*.
5. **Pengujian Sistem dan Evaluasi**, tahap pengujian dilakukan untuk memastikan sistem sudah berjalan sesuai dengan yang dibutuhkan. Selanjutnya proses evaluasi dilakukan dengan menggunakan *confusion matrix* untuk menghitung hasil performa dari *accuracy*, *precision*, *recall* dan *f1-score*.
6. **Konsultasi Penelitian dan Laporan**, pada tahap ini, penulis melakukan bimbingan kepada dosen pembimbing agar penelitian yang telah dibuat berjalan dengan sempurna dan mendokumentasikan hasil penelitian ke dalam bentuk laporan agar penelitian ini dapat dikembangkan lebih lanjut kedepannya.

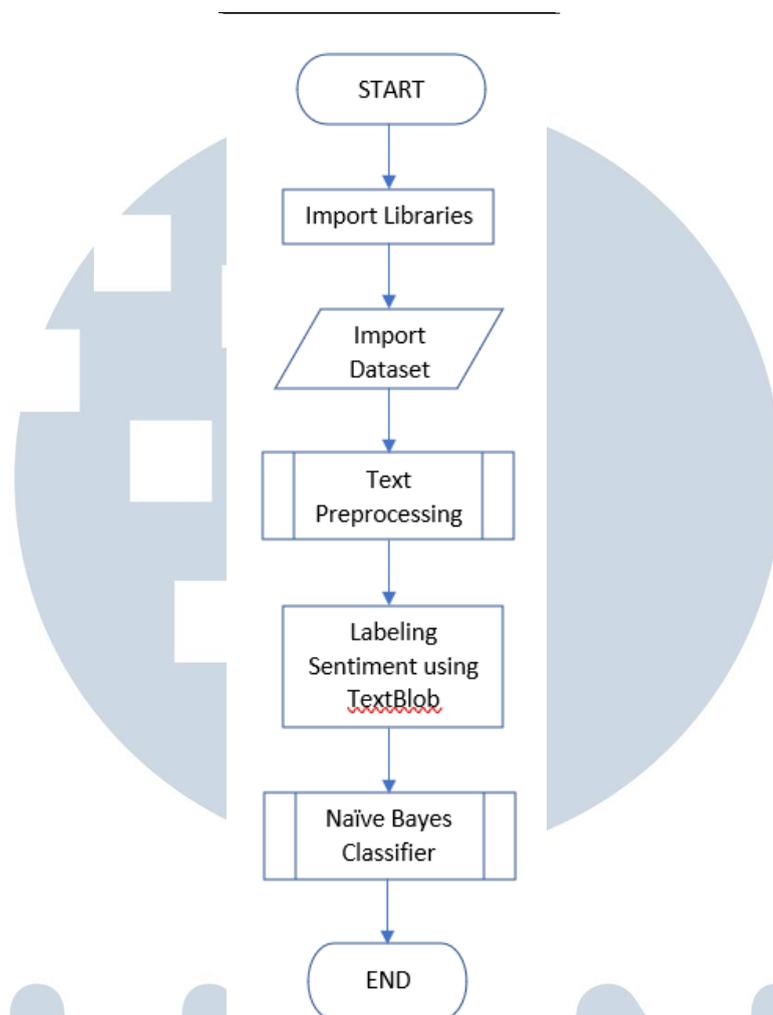
3.2 Perancangan Sistem

Perancangan sistem implementasi algoritma *naive bayes* untuk analisis sentimen review produk di Tokopedia akan dijabarkan dalam *flowchart* yang terdiri dari *flowchart* utama, *flowchart preprocessing*, dan *flowchart naive bayes*.

3.2.1 *Flowchart* Utama

Gambar 3.1 merupakan *flowchart* utama pada aplikasi analisis sentimen menggunakan naive bayes. Proses tahap pertama adalah *import library* yang dibutuhkan dalam penelitian. Lalu proses selanjutnya adalah *import* dataset dengan file berbentuk .csv. Proses selanjutnya adalah tahap *preprocessing*. *Preprocessing* merupakan proses untuk seleksi data teks agar menjadi kalimat yang terstruktur dan efisien. Dalam *preproceesseing* terdiri dari beberapa tahap, yaitu *cleaning text*, *case folding*, *tokenization*, *stopword*, dan *stemming*. Lalu proses selanjutnya adalah melakukan labeling sentimen pada tiap review dari pengguna. Labeling sentimen terbagi menjadi 3, yaitu positif, negatif, dan netral. Untuk melakukan labeling sentimen dilakukan secara otomatis dengan menggunakan library Text Blob. Proses selanjutnya adalah melakukan klasifikasi menggunakan algoritma Naive Bayes. Naive Bayes digunakan untuk mencari atau memprediksi nilai *probability* dari tiap klasifikasi.





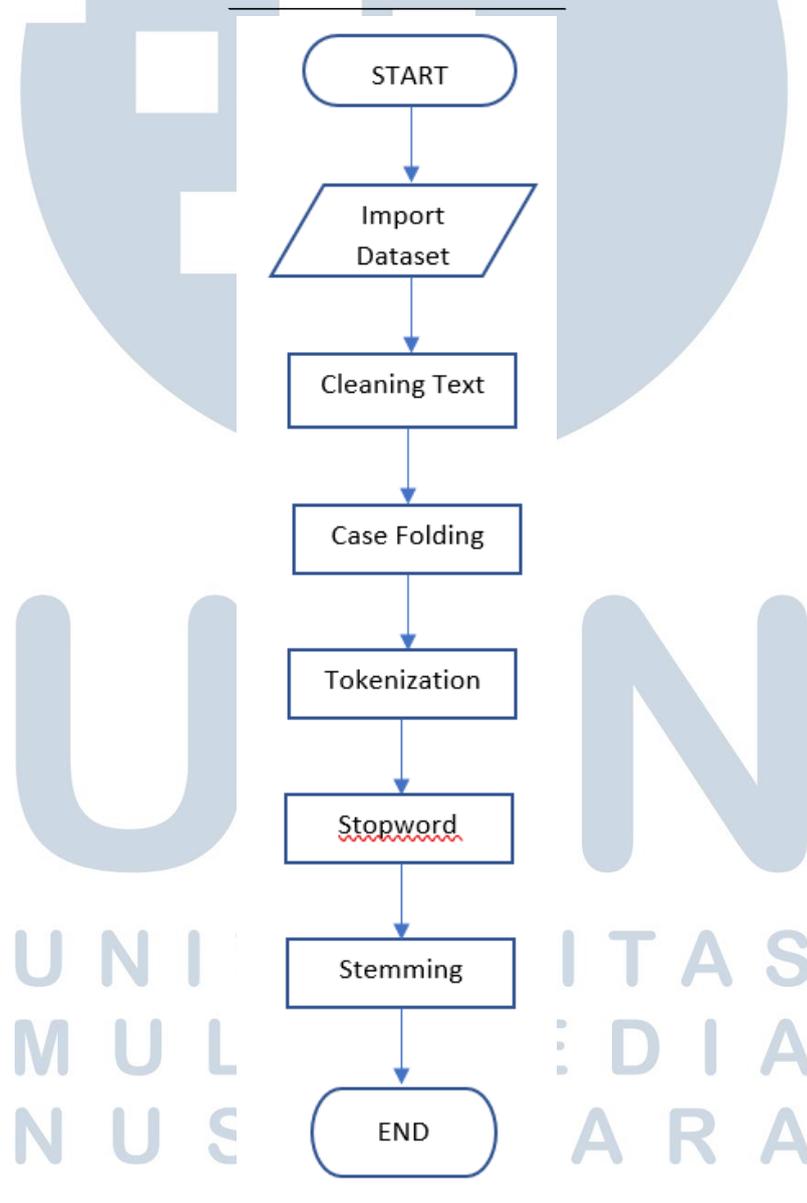
Gambar 3.1. Flowchart Utama

3.2.2 *Flowchart Preprocessing*

Gambar 3.2 merupakan tahap *preprocessing*, yaitu proses pemilihan teks agar hasil teks tersebut menjadi lebih terstruktur dan efisien. setelah dataset diimport, dilakukan proses preprocessing. Terdapat beberapa tahapan pada proses Preprocessing, adalah sebagai berikut:

1. *Cleaning Text*, merupakan proses untuk membersihkan data teks dari simbol-simbol dan karakter tanda baca yang tidak perlu digunakan.
2. *Case Folding*, merupakan proses untuk merubah semua huruf pada kalimat menjadi huruf kecil.

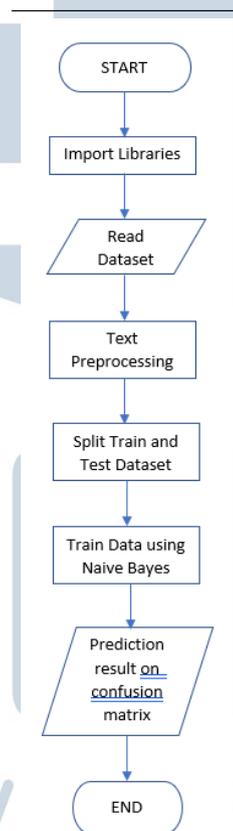
3. *Tokenization*, merupakan proses untuk memecah sekumpulan karakter pada teks menjadi satuan kata.
4. *Stopword*, merupakan proses untuk menghilangkan kata umum yang sering muncul dan tidak memiliki arti.
5. *Stemming*, merupakan proses untuk mereduksi bentuk kata menjadi kata dasar.



Gambar 3.2. Flowchart Preprocessing

3.2.3 Flowchart Model Klasifikasi Naive Bayes

Gambar 3.3 merupakan *flowchart* atau alur proses dari klasifikasi Naive Bayes. Setelah dataset diproses pada *text preprocessing* untuk menghasilkan kalimat yang terstruktur dan efisien, dibagi menjadi 2 bagian, yaitu *data train* dan *data test*. Terdapat 2 skenario dalam melakukan proses *split train* dan *test*, yaitu dengan menggunakan rasio perbandingan 70:30 dan rasio perbandingan 80:20. *Data train* akan *training* atau dilatih, dan Naive Bayes melakukan prediksi pada *data test* dengan mencocokkan hasil *X test* dan *y test* untuk mendapatkan nilai dari *accuracy*, *precision*, *recall*, dan *f1-score* dari model pengelompokan tabel *confusion matrix*.



Gambar 3.3. Flowchart Naive Bayes