BAB 3
METODE PENELITIAN

Penelitian ini merupakan studi analisis sentimen yang memfokuskan pada aspek positif dan negatif dari ulasan pelanggan Maula Hijab di Tokopedia. Metode yang digunakan yaitu Naive Bayes. Penelitian memerlukan data ulasan yang diambil dari Tokopedia Maula Hijab. Selanjutnya, melakukan proses *preprocessing* untuk memperoleh data yang sesuai dengan kebutuhan. Data-data tersebut kemudian digunakan dalam menganalisis sentimen atau pendapat dari pelanggan Tokopedia mengenai produk Maula Hijab.

Penelitian dimulai dengan mengidentifikasi latar belakang permasalahan, melakukan pengolahan data yang ada, membentuk model topik, dan mencari nilai-nilai sentimen yang optimal untuk memperoleh informasi sesuai dengan tujuan yang diinginkan (Putra Aziziya et al., 2022). Berikut ini adalah komponen-komponen penelitian analisis sentimen produk Maula Hijab di Tokopedia beserta dengan langkah-langkah yang diambil dalam proses analisis sentimen menggunakan data ulasan.

3.1 BAHAN PENELITIAN

Bahan penelitian yang diperlukan untuk penelitian ini adalah ulasan dan rating yang diberikan oleh pelanggan Maula Hijab di platform Tokopedia. Data ini mencakup teks ulasan yang ditulis oleh pelanggan bersama dengan penilaian (rating) yang mereka berikan. Ulasan ini dapat berisi pendapat, pengalaman, dan *feedback* terkait produk atau layanan yang diberikan oleh Maula Hijab.

3.2 ALAT PENELITIAN

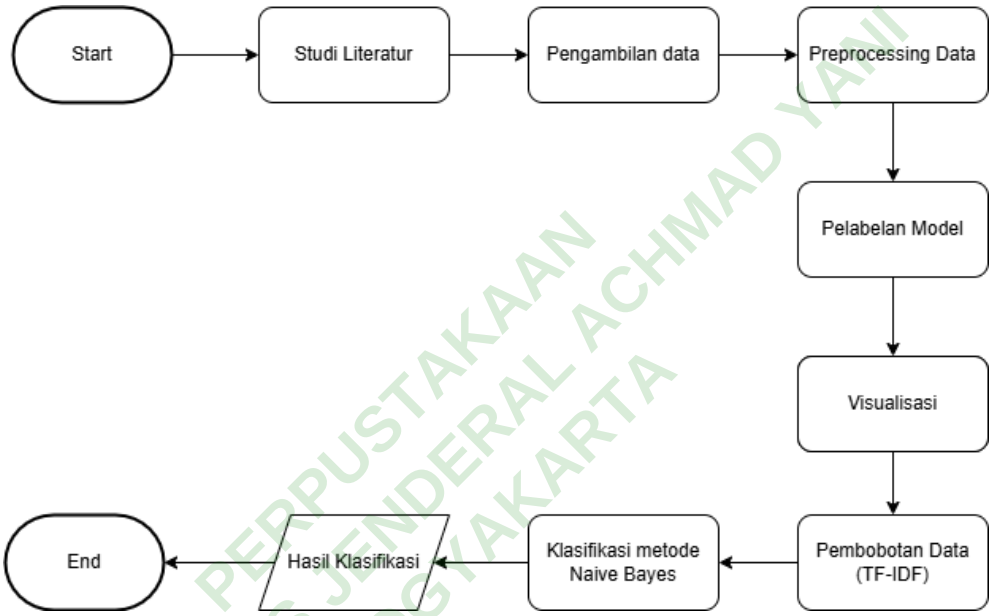
Sistem Operasi dan program-program aplikasi yang dipergunakan dalam dalam pengembangan penelitian ini adalah:

1. Sistem Operasi: Windows 10 atau yang lebih baru.
2. *Website* Tokopedia
3. *Microsoft Excel*

4. *Visual Studio Code*
5. Pemrograman python

3.3 JALAN PENELITIAN

Dalam jalan peneltian atau alur yang diperlukan, ialah sebagai berikut menggunakan beberapa jenis/metode yaitu:



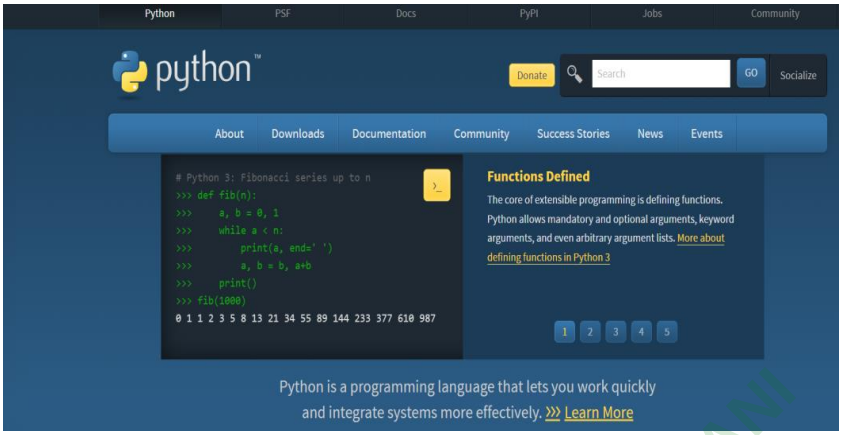
Gambar 3.1 Diagram Jalan Penelitian.

1. Studi Literatur

   Pada metode ni penulis akan membaca jurnal atau buku dari sumber yang berkaitan atau berhubungan dengan penulisan tugas akhir.

2. Pengambilan Data

   Pada tahap pertama dijelaskan di gambar 3.2 akan menggunakan bahasa pemrograman Python untuk proses pengumpulan data. *Library* yang akan ditambahkan ialah, *Selenium*, *Time*, dan *Csv*.

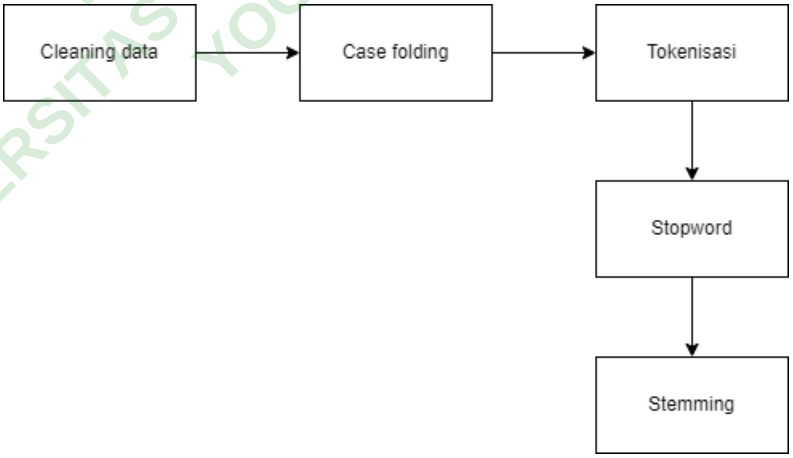


Gambar 3.2 Bahasa Pemrograman Python

Pada tahap pertama, dijelaskan pada gambar 3.3 akan menggunakan bahasa pemrograman Python. Pengambilan data ini menggunakan *library Selenium*, *csv*, dan *time.*

3. *Preprocessing* Data.

   Pada proses *Preprocessing* data untuk melakukan pengolahan data tersebut dalam pemrograman Python. *Library* yang akan di tambahkan ialah Pandas, *Re, String, Nltk* dan *Sastrawi*.



Gambar 3.3 Alur *Preprocessing* data.

Gambar 3.3 di atas bertujuan untuk membersihkan sebuah kata pada teks. Berikut penjelasan mengenai alur *preprocessing* data :

a. *Cleaning* data

Tahap pertama dalam *preprocessing* data ialah *cleaning* data. *Cleaning* berfungsi untuk menghilangkan tanda baca pada teks di suatu kalimat, karakter khusus, spasi ganda, dan elemen non-teks lainnya seperti HTML tags atau emojis. Contoh : “Kamu tanya?” menjadi “Kamu tanya”. (Kusuma H.W, 2023).

b. *Case folding*

Setelah melakukan tahapan *cleaning*, selanjutnya adalah *case folding*. Fungsi ini ialah mengubah huruf dalam teks menjadi huruf kecil atau huruf besar menjadi senada. Contoh : “Belanja Online” atau “BELANJA ONLINE” menjadi “belanja online”. Tahap ini berfungsi agar tidak terjadinya duplikasi teks atau kata (Merinda Lestandy et al., 2021)

c. Tokenisasi

Setelah melakukan tahapan *case folding* , selanjutnya ialah memecahkan teks menjadi kata-kata (token). Contoh : “Saya suka berbelanja” menjadi [“Saya”, “suka”, “berbelanja”] (Merinda Lestandy et al., 2021).

d. *Stopword*

Setelah melakukan tahapan tokenisasi tahap selanjutnya, ialah menghapus kata-kata umum yang tidak memberikan informasi penting. Contoh : “dan”, “di”, “yang”.

e. *Stemming*

Setelah melakukan tahapan *stopword*, selanjutnya ialah mengubah kata menjadi (kata dasar). Contoh: “berbelanja” menjadi “belanja” (Amrullah et al., 2020)

4. Pelabelan Model.

Proses ini akan mengkategorikan sebuah teks. Apakah teks tersebut bersifat positif atau negatif. Proses ini akan dilakukan secara manual .

5. Visualisasi.

   Selanjutnya, ialah tahap visualisasi di tahap ini bertujuan untuk menampilkan frekuensi kata dalam suatu dokumen atau kumpulan teks. Kata yang muncul akan berukuran besar, sedang, atau kecil, semakin sering kata tersebut muncul akan menghasilkan *output* semakin besar pada teks atau kata tersebut.(Josen Limbong et al., n.d.)

6. Pembobotan Data.

   Proses pembobotan data ini menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*). Fungsi penggunaan TF-IDF dalam pembobotan data ialah untuk mengukur pentingnya suatu kata dalam sebuah data.

7. Klasifikasi metode Naïve Bayes.

   Pada tahap ini, data yang telah diolah akan digunakan untuk melatih model Naïve Bayes. Metode ini dipilih karena kesederhanaan dan efektivitasnya dalam klasifikasi teks. Model ini akan mempelajari hubungan antara kata-kata (fitur) dengan sentimen (positif atau negatif) berdasarkan data pelatihan yang telah dilabeli sebelumnya.

8. Hasil Klasifikasi.

   Setelah model Naïve Bayes selesai dilatih, model tersebut akan digunakan untuk mengklasifikasikan teks-teks baru yang belum diketahui sentimennya. Setiap teks akan dianalisis berdasarkan kata-kata yang terkandung di dalamnya, dan model akan memberikan prediksi sentimen positif atau negatif untuk teks tersebut. Untuk mengevaluasi kinerja model, hasil prediksi akan dianalisis menggunakan confusion matrix. Output yang di dapatkan ialah akurasi, *precision, recall, dan f1-score*. Akurasi akan mengindikasikan seberapa baik model dalam mengklasifikasikan teks-teks baru berdasarkan data uji.