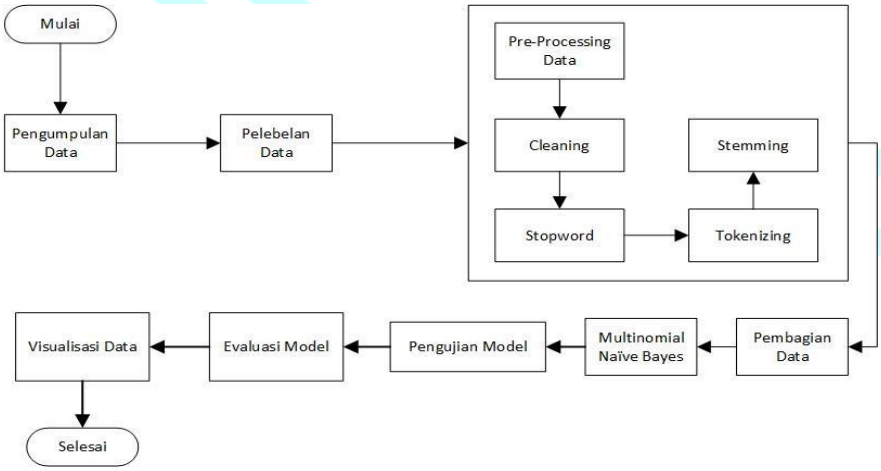
BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Penelitian ini berfokus pada ulasan pengguna bank digital pada bulan April 2025 aplikasi Seabank, Blu by BCA, Krom Bank, Bank Saqu, Bank Jago. Data didapatkan dari ulasan pengguna di *Google Play Store*, mencakup komentar berbentuk teks yang memuat opini pengguna, seperti pujian, kritik, dan masukan kepada pemilik aplikasi untuk melakukan perbaikan di masa mendatang. Analisis terhadap ulasan tersebut dilakukan menggunakan algoritma *Naive Bayes*, yang berfungsi untuk mengklasifikasikan sentimen pengguna ke dalam dua kategori yaitu positif dan negatif.

3.2 Tahapan Penelitian

Penelitian ini melibatkan beberapa tahapan utama yang disusun secara terstruktur, sebagaimana ditampilkan dalam diagram alur (*flowchart*) di bawah. Setiap tahap diuraikan secara mendetail untuk memberikan pemahaman menyeluruh mengenai keseluruhan proses penelitian. Dalam penelitian ini, penulis melakukan sentimen data ulasan pengguna untuk mengklasifikasikan data ulasan ke dalam kategori positif dan negatif menggunakan metode *Naive Bayes*. Adapun langkah-langkah dalam melakukan klasifikasi data terdapat pada gambar di bawah.



Gambar 3. 1 Tahapan penelitian

3.3 Pengumpulan Data

Pengumpulan data Merupakan tahap pengumpulan data yang diperlukan untuk menyelesaikan suatu permasalahan tertentu(Tanjung et al., 2023),(Sfofiah Hilabi et al., n.d.). Data didapatkan untuk memperoleh ulasan pengguna aplikasi Seabank, Blu by BCA, Krom Bank, Bank Saqu, Bank Jago yang tersedia di *Google Play Store* menggunakan metode *Web Scraping*. Proses ini dilakukan untuk mempersiapkan data sebelum dianalisis lebih lanjut menggunakan algoritma *Naive Bayes*.

3.4 Pelebelan Data

Pelebelan data dilakukan setelah data terkumpul untuk mempermudah analisis sentimen. Pelebelan data adalah proses pemberian kategori sentiment seperti positif, negatif, atau netral pada setiap komentar(Alvionika et al., 2024),(Chusyairi, n.d.).

3.5 *Pre-Processing Data*

Proses selanjurnya adalah *Pre-processing* data, tahapan ini dilakukan untuk membersihkan data sebelum data di uji dengan *Naive Bayes. Pre-processing* ini meliputi tahapan diantaranya *cleaning*, *stopword* , *tokenizing*, dan *stemming*.

3.5.1 *Cleaning*

Pada tahap *cleaning*, karakter yang tidak penting atau tidak relevan seperti tanda baca, angka, atau simbol khusus dapat dihapus(Prakoso & Hermawan, 2023). Pada tahap *cleaning* ini membersihkan teks dari tanda baca ataupun karakter khusus yang tidak diperlukan dan mempertahankan informasi yang relevan pada ulasan.

3.5.2 *Stopword*

Tahapan ini untuk menghilangkan kata tidak penting dan tidak memiliki arti seperti kata “dan”, “yang”, ”atau”, filtering juga dapat mengurangi dimensi terhdap data *text input*, sehingga proses akan lebih mudah dijalankan(Selawati et al., 2025).

3.5.3 *Tokenizing*

Tokenizing atau tokenisasi merupakan proses pembagian teks berupa paragraf atau kalimat menjadi beberapa bagian. Tokenisasi ini menggunakan

delimiter atau pemisah pada setiap kata yaitu berupa spasi, enter, atau whitespace(Lustiansyah et al., n.d.),(Suarna & Prihartono, 2024).

3.5.4 *Stemming*

Proses mengurangi kata-kata ke bentuk dasarnya dengan menghapus imbuhan(Nur Oktavia et al., 2024). Salah satu algoritma yang dipakai untuk menghapus imbuhan dalam adalah *Stemming* Sastrawi. Sastrawi adalah algoritma stemmer yang digunakan untuk mengatasi tantangan mengubah kata-kata menjadi kata-kata sederhana(Carolina et al., n.d.),(Wicaksono & Cahyono, 2024).

3.6 Pembagian Data

Pembagian data atau *splitting data* merupakan proses pembagian data yang digunakan dalam penelitian menjadi dua atau lebih bagian, biasanya digunakan untuk menguji model atau algoritma. Dataset umumnya dibagi menjadi data(training) dan data uji(testing)(Putri et al., 2023).

3.7 *Multinomial Naive Bayes*

Naive Bayes merupakan metode yang dapat memprediksi pada suatu class dari suatu objek yang mana kelasnya tidak diketahui dari masing – masing kelompok atribut yang ada sehingga penentuan class yang paling optimal dapat didasarkan pada pengaruh didapat di hasil pengamatan(Nuranisah & Yanti Yusman, 2023). *Multinomial Naive Bayes* (MNB) adalah metode klasifikasi khusus untuk data teks. Keunggulannya terletak pada pendekatan independen, di mana setiap dokumen diklasifikasikan secara terpisah berdasarkan isinya tanpa dipengaruhi oleh dokumen lain(Elwanda Putra & Frazna Azzahra, 2025),(Huda et al., n.d.)

3.8 Pengujian Model

Untuk mengetahui seberapa akurat prediksi yang dibuat, tahap pengujian model dilakukan untuk mengevaluasi kinerja model yang telah dibangun dengan data uji. Nilai akurasi menunjukkan seberapa jauh nilai sebenarnya dan nilai prediksi berbeda. Akurasi didefinisikan sebagai tingkat keselarasan antara informasi yang dikehendaki serta respon yang diberikan sistem(Lestari et al., 2023).

3.9 Evaluasi Model

Evaluasi model merupakan tahap krusial dalam pengembangan dan penerapan model, yang bertujuan mengukur dan menilai performa model berdasarkan hasil pengujian terhadap data yang telah disiapkan. Salah satu metode evaluasi yang umum digunakan adalah *confusion matrix*, yang menjadi dasar dalam menghitung metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*, dengan demikian dapat menunjukkan seberapa baik model dalam melakukan klasifikasi terhadap data(Anggi Pangesti et al., n.d.),(Tukino et al., 2024). Berikut merupakan sejumlah formula evaluasi untuk menilai kinerja model klasifikasi, dengan cara membandingkan hasil prediksi model terhadap data sebenarnya dalam proses pengklasifikasian.

3.9.1 *Accuracy*

Menilai proporsi prediksi yang tepat dibandingkan dengan total keseluruhan data yang diuji.

$$accuracy = \frac{TP+TN}{Total} \quad (1)$$

3.9.2 *Precision*

Mengukur ketepatan prediksi positif, yaitu prediksi positif yang benar positif.

$$precision = \frac{TP}{TP+FP} \quad (2)$$

3.9.3 *Recall*

Menilai sejauh mana model mampu mendeteksi seluruh kasus positif yang benar-benar ada.

$$recall = \frac{TP}{TP+FN} \quad (3)$$

3.9.4 *F1-Score*

Merupakan nilai rata-rata harmonik dari *precision* dan *recall* yang berfungsi untuk menyeimbangkan kontribusi keduanya dalam evaluasi performa model.

$$F1\ Score = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (4)$$

Keterangan:

TP (*True Positive*): Kasus positif yang berhasil diprediksi dengan benar sebagai positif.

TN (*True Negative*): Kasus negatif yang berhasil diprediksi dengan benar sebagai negatif.

FP (*False Positive*): Kasus negatif yang keliru diprediksi sebagai positif.

FN (*False Negative*): Kasus positif yang keliru diprediksi sebagai negatif.

3.10 Visualisasi Hasil

Visualisasi hasil menampilkan data dan hasil analisis dalam bentuk grafis untuk memudahkan pemahaman dalam mempresentasikan hasil data. Pada tahap visualisasi hasil ini, penulis akan menggunakan diagram batang yang menunjukkan presentase sentimen data ulasan positif ataupun negatif pada aplikasi bank digital.

UBP KARAWANG