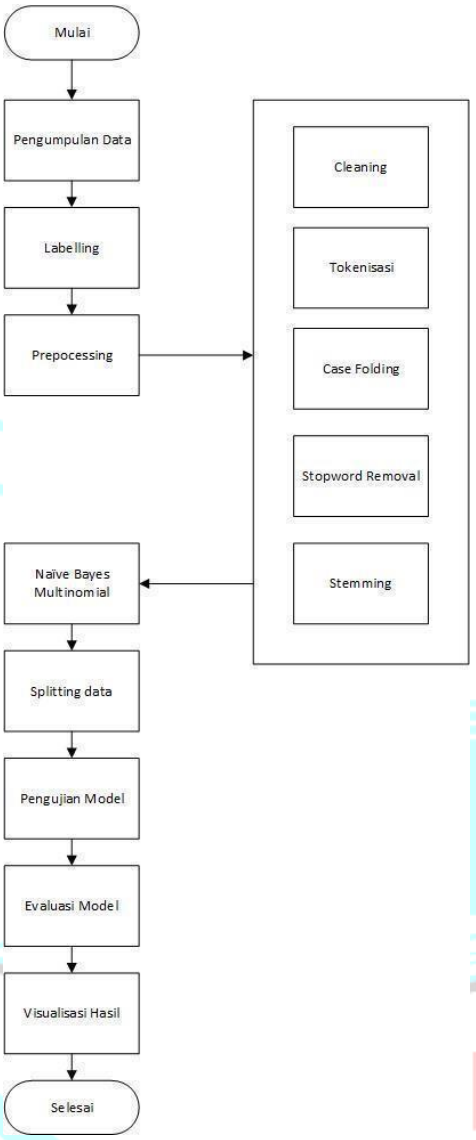
BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Penelitian ini berfokus pada data ulasan pengguna dari sebuah aplikasi Amazon Prime Video, Disney+Hotstar, CatchPlay+, Netflix, dan Viu dari bulan Maret 2025. Ulasan pengguna di Google Play Store menjadi sumber data utama dalam penelitian ini. Data tersebut mencakup ulasan komentar pengguna yang menggambarkan tingkat kepuasan pengguna secara umum, serta komentar teks yang berisi opini pengguna, seperti pujian, keluhan, dan saran untuk pengembangan aplikasi. Analisis terhadap data ulasan ini dilakukan dengan menggunakan algoritma Naive Bayes untuk mengelompokkan sentimen pengguna ke dalam kategori positif dan negatif.

3.2 Prosedur 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, Netral dan Negatif menggunakan metode Naive Bayess. Adapaun langkah-langkah dalam melakukan klasifikasi data terdapat pada gambar di bawah.



Gambar 3. 1 Metode Penelitian

3.2.1 Pengumpulan Data

Penelitian melibatkan pengumpulan informasi atau fakta penting untuk menjawab pertanyaan, menjadikan pengumpulan data tahap penting(Paryono et al., 2023). Studi ilmiah, survei, analisis data, atau pengembangan produk adalah beberapa contoh bidang di mana hal ini dapat dilakukan (Rifa'i, n.d.).

3.2.2 Labelling

Labelling adalah proses memberikan kategori atau kelas pada data, misalnya sentimen positif dan negatif, untuk digunakan dalam pelatihan dan pengujian model machine learning seperti Naive Bayes. Dalam analisis sentimen, labelling bertujuan untuk menyediakan data yang sudah ditandai (labeled data) sehingga algoritma dapat mempelajari pola-pola yang ada untuk mengklasifikasikan teks baru(Wiranata et al., 2023),(Shofa Shofiah Hilabi, 2021).

3.2.3 Prepocessing

Data preprocessing adalah langkah kritis dalam pipeline data science yang dapat mempengaruhi hasil akhir analisis secara signifikan(Hamdani et al., 2024),(Lia Hananto et al., 2021). Langkah-langkah umum dalam preprocessing mencakup cleaning, tokenisasi, case folding, stopword removal, dan stemming. Tahap ini penting untuk meningkatkan akurasi model analisis sentiment. Adapun taapan preprocessing sebagai berikut:

1. Cleaning : Proses cleaning data melibatkan penghilangan tanda baca, angka, dan tautan yang tidak relevan(Andini et al., 2022)
2. Tokenisasi :Tokenisasi dilakukan dengan memecah teks menjadi kata-kata atau frasa. Proses ini penting dalam pemrosesan bahasa alami karena memungkinkan analisis lebih lanjut terhadap struktur dan makna teks(Lutfiyani & Retnowati, 2021)
3. Case Folding : Case folding menyamakan format huruf dengan mengubah semua huruf menjadi kecil(Cahya Kamilla et al., 2024).
4. Stopword Removal : Stopword removal mengeliminasi kata-kata yang tidak memberikan informasi signifikan, seperti "dan", "atau", "yang", dan lain-lain. Penghapusan stopword bertujuan untuk mengurangi noise dalam data dan meningkatkan efisiensi serta akurasi dalam proses klasifikasi teks(Irfan & Erizal, 2024)
5. Stemming : Mengubah kata menjadi bentuk aslinya disebut stemming. Menyeragamkan bentuk kata dilakukan melalui stemming. Stemming bertujuan untuk mengurangi kata ke bentuk dasarnya(Yuniar et al., 2022).

3.2.4 Naive Bayes Multinomial

Model algoritma Naive Bayes Multinomial merupakan salah satu metode pembelajaran berbasis probabilistik yang sering diterapkan dalam analisis bahasa alami (NLP). Prinsip dasar algoritma ini mengandalkan teorema Bayes dengan memanfaatkan frekuensi kata dalam dokumen sebagai variabel utama (Yuyun et al., 2021),(Priyatna & Novalia, 2023).

3.2.5 Spliting Data

Dataset biasanya dibagi menjadi data (latih) dan data uji (pengujian). Splitting data adalah proses pembagian data yang digunakan dalam penelitian menjadi dua atau lebih bagian (Putri et al., 2023).

3.2.6 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),(Kulsum et al., 2022).

3.2.7 Evaluasi Model

Tahap penting dalam proses pengembangan model pembelajaran mesin adalah evaluasi model untuk menilai kinerjanya. Metode evaluasi yang umum digunakan tercantum akurasi, presisi, recall, dan skor F1 (Khoirul et al., 2023). Evaluasi yang efektif membantu dalam memilih model terbaik dan menemukan masalah yang perlu diperbaiki.

$$Akurasi = \frac{TP + TN}{\text{Total}}$$

$$Precision = \frac{TP}{\text{TP} + \text{FP}}$$

$$Recall = \frac{TP}{\text{TP} + \text{FN}}$$

$$f1\,Score = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Keterangan:

TP (True Positive): Jumlah data positif yang diklasifikasikan benar oleh model.

TN (True Negative): Jumlah data negatif yang diklasifikasikan benar oleh model.

FP (False Positive): Jumlah data negatif yang salah diklasifikasikan sebagai positif oleh model.

FN (False Negative): Jumlah data positif yang salah diklasifikasikan sebagai negatif oleh model.

3.2.8 Visualisasi Hasil

Visualisasi hasil adalah proses menyajikan data dan hasil analisis dalam bentuk grafis untuk memudahkan interpretasi dan pemahaman. Dalam analisis sentimen, visualisasi dapat berupa grafik batang, pie chart, atau word cloud yang menunjukkan distribusi sentimen atau kata-kata yang sering muncul (Arsi & Waluyo, 2021),(Agustia Hananto et al., 2024).