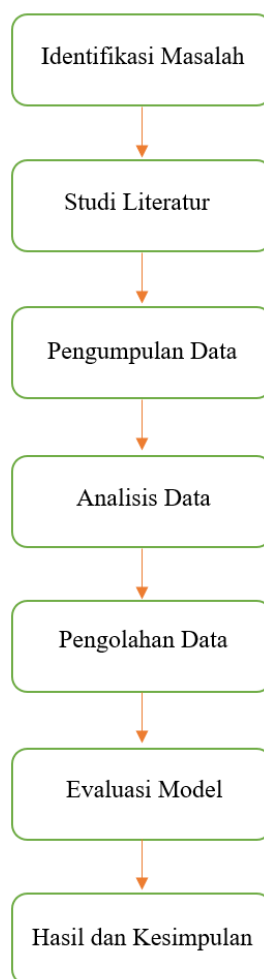


BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Desain ini mencakup algoritma yang dipilih, sumber data, strategi analisis, dan tahapan pelaporan, guna memastikan bahwa teknik yang diterapkan bisa memastikan data yang diperoleh memiliki validitas tinggi serta mendukung tujuan penelitian. Berikut merupakan tampilan dari Desain Penelitian:



Gambar 3.1 Desain Penelitian

Berikut merupakan penjabaran dari desain penelitian:

1. Identifikasi Masalah

Mengklasifikasikan opini penonton *film* pada *platform streaming video* merupakan masalah yang diidentifikasi dalam penelitian ini.

2. Studi Literatur

Penelitian ini didukung oleh studi literatur yang terdiri dari studi ilmiah atau penelitian terdahulu yang membahas penerapan pendekatan *naïve bayes*. Dalam studi ini akan membahas tentang konsep, metode, hasil, dan kesimpulan yang berkaitan dengan penerapan pendekatan *naïve bayes* untuk mengklasifikasi opini penonton *film* berdasarkan sentimen positive, negative, dan netral.

3. Pengumpulan Data

Pengumpulan data yang dimaksud melibatkan pengumpulan informasi atau data yang relevan untuk mengkategorikan opini penonton berdasarkan tingkat kepuasan mereka terhadap layanan *streaming video*. Salah satu metode untuk mendapatkan data komentar dari evaluasi aplikasi *Netflix* dan *Disney+* di *Google Play Store* adalah *web scraping*.

4. Analisis Data

Evaluasi informasi yang dikumpulkan tentang opini penonton dikenal sebagai analisis data, dan dapat digunakan untuk menilai seberapa jauh kualitas tontonan berdasarkan kebutuhan penonton.

5. Pengolahan Data

Setelah data dianalisis, data akan dibersihkan melalui *Google Colabs* dan hasilnya akan dihitung menggunakan teknik data mining *Naive Bayes*. Pada langkah ini, penghapusan duplikat, pengangan nilai hilang (*NaN*), koreksi kesalahan ketik, teks normalisasi (mengubah tulisan jadi huruf kecil dan menghapus tanda baca), dan teks ditokenisasi menjadi unit-unit kecil (kata atau frasa). Setelah pembersihan data, ekstraksi fitur dilakukan menggunakan metode seperti *Bag-of-Words (BoW)* atau *TF-IDF (Term Frequency-Inverse Document Frequency)* dalam mengganti teks ke model numerik (angka) agar bisa dimengerti algoritma.

6. Evaluasi Model

Setelah model dilatih, dengan mengevaluasi seberapa baik kinerja model *Naive Bayes* dalam menghasilkan prediksi pada data sebelumnya yang tidak diketahui. Tahapan pada pendekatan ini yang pertama yaitu menggunakan model untuk memperkirakan sentimen data uji (X_{test}). Sentimen aktual (y_{test}), yang sebelumnya diketahui, kemudian dibandingkan dengan hasil prediksi (y_{pred}). Kinerja model dinilai menggunakan berbagai metrik evaluasi, akurasi menunjukkan visualisasi jumlah perkiraan yang akurat dan tidak akurat pada masing-masing kelas, laporan klasifikasi (presisi, *recall*, dan skor F1) menunjukkan kinerja terperinci untuk setiap kategori sentimen dan akurasi memberikan ringkasan persentase prediksi yang benar.

7. Hasil dan Kesimpulan

Hasil dan kesimpulan bertujuan untuk memberikan pemahaman menyeluruh tentang sejauh mana hasil pengujian dapat memberikan data yang relevan dan platform dapat meningkatkan layanan di masa mendatang dengan kualitas yang lebih baik.

3.2 Variabel

Aspek atau faktor tertentu yang diukur atau teliti dalam sebuah penelitian merupakan variabel penelitian. Berikut merupakan beberapa variabel penelitian:

1. Variabel Independen (Fitur Input)

Variabel ini terdiri dari teks ulasan mentah, teks ulasan yang telah dibersihkan, dan fitur ekstraksi (vektor numerik) yang akan kita gunakan untuk memprediksi sentimen.

2. Variabel Dependen (Label Sentimen)

Variabel ini menampilkan sentimen sebuah komentar positif: Menunjukkan kepuasan atau pendapat yang baik, Negatif: Menunjukkan ketidakpuasan atau pendapat negatif, Netral: Tidak mengungkapkan pendapat yang jelas atau akurat.

3. Variabel Evaluasi (Metrik Kinerja Model)

Akurasi, presisi, recall (*Sensitivity* atau *True Positive Rate*), Skor F1, *Confusion Matrix*, dan Kurva *AUC-ROC* (*Area Under the Receiver Operating Characteristic Curve*) untuk menilai seberapa baik model dalam memprediksi sentiment merupakan metrik yang diterapkan.

3.3 Populasi Dan Sampel

3.3.1 Populasi

Pengguna layanan *streaming video Netflix* dan *Disney+ Hotstar* merupakan bagian dari populasi penelitian ini. Sebanyak 1.000 data ulasan yang dikumpulkan di aplikasi *Google Play Store* akan digunakan untuk penilaian pengguna dari kedua *platform* tersebut.

3.3.2 Sampel

Untuk menjamin bahwa sentimen (positif, negatif, dan netral) direpresentasikan secara proporsional pada ulasan pengguna, peneliti menerapkan *Stratified Random Sampling*. Dari setiap aplikasi, penulis mengambil 1.000 ulasan secara acak, sehingga total ukuran sampel menjadi 2.000 ulasan. Tujuan dari pilihan ini adalah untuk memastikan bahwa representasi kedua layanan *streaming video* dilakukan secara seimbang.

3.4 Teknik Pengumpulan Data

Tahapan pengumpulan data atau informasi berdasar sumber tertentu dikenal sebagai teknik pengumpulan data. Yang bertujuan guna menghasilkan data yang lebih akurat, relevan, dan dapat dipercaya untuk analisis lebih lanjut. Berikut ini adalah teknik pengumpulan data yang digunakan dalam penelitian ini:

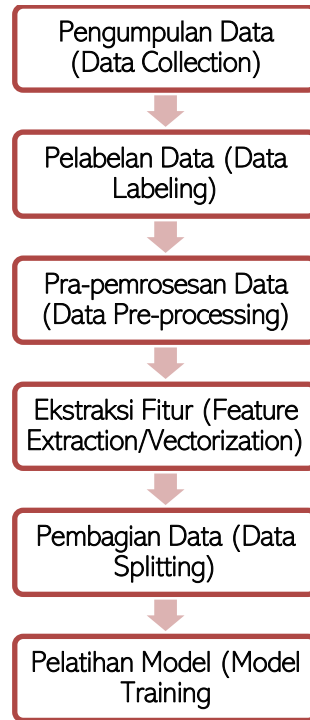
1. Teknik Dokumentasi: opini yang berasal dari pengguna *Netflix* dan *Disney+* pada *platform Google Play Store* diambil untuk pengumpulan data. Data opini pengguna secara langsung dan autentik dari pengalaman

penggunaan aplikasi streaming video diperoleh menggunakan teknik ini. Komentar-komentar tersebut berisi opini, keluhan, atau pujian pengguna tentang fitur, kualitas layanan, aksesibilitas konten, dan fungsionalitas aplikasi. Dengan menggunakan bantuan alat (*tools*) atau pustaka (*library*) *web scraping* seperti *Google Colab* atau *Python*, data komentar dikumpulkan. Algoritma *Naive Bayes* untuk klasifikasi sentimen kemudian dilatih dan diuji menggunakan data komentar ini sebagai dataset.

2. Untuk memperkuat kerangka teori dan pilihan teknik klasifikasi, studi literatur perlu dilakukan. Seperti jurnal ilmiah, publikasi, dan laporan relevan terkait tentang analisis sentimen dalam domain aplikasi streaming, performa algoritma *Naive Bayes* dalam klasifikasi teks, metode *preprocessing*, dan ekstraksi fitur teks disertakan dalam literatur yang diteliti. Meningkatkan pengetahuan dan landasan teori praktik terbaik dalam pengolahan opini digital merupakan maksud dari riset ini.

3.5 Metode Analisis Data

Berikut merupakan *flowchart* untuk metode analisis data:



Gambar 3.2 Metode Analisis Data

3.6 Lokasi Dan Jadwal Penelitian

3.6.1 Lokasi

Lokasi studi ini berada pada teritori Indonesia, salah satu indikator paling akurat berdasarkan komentar atau ulasan pengguna *platform streaming video Netflix* dan *Disney+* pada aplikasi *Google Play Store* mengenai reaksi, opini, dan tingkat kepuasan pengguna terhadap *platform* ini.

3.6.2 Jadwal Penelitian

Berdasarkan tabel di bawah ini menampilkan rentang waktu penelitian selama 4 bulan untuk mengklasifikasikan opini penonton film, yang berlangsung dari April 2025 hingga Juli 2025. Terdapat 6 aktivitas yang dilakukan yakni, identifikasi masalah, pengumpulan data, analisis data, pengelolaan data, evaluasi model serta hasil merupakan aktivitas yang akan diselesaikan selama kurun waktu tersebut. Berikut merupakan jadwal dari penelitian yang akan di lakukan:

Tabel 3.1 Jadwal Penelitian

[illegible]