

BAB 3

METODE PENELITIAN

Studi ini mengangkat penerapan *Data Mining*, dimulai dengan menguraikan latar belakang fenomena masalah yang terjadi dan melakukan pencarian data yang akan diproses untuk menciptakan klasifikasi persentase dari setiap tanggapan pengguna dalam sentimen positif maupun negatif. Berbagai sumber dan perangkat digunakan dalam penelitian ini untuk menilai tingkat kepuasan pelanggan berdasarkan rating maupun opini yang disampaikan oleh pengguna mengenai aplikasi Vidio dan Loklok pada platform Google Play Store dengan menerapkan metode Naïve Bayes.

3.1 BAHAN DAN ALAT PENELITIAN

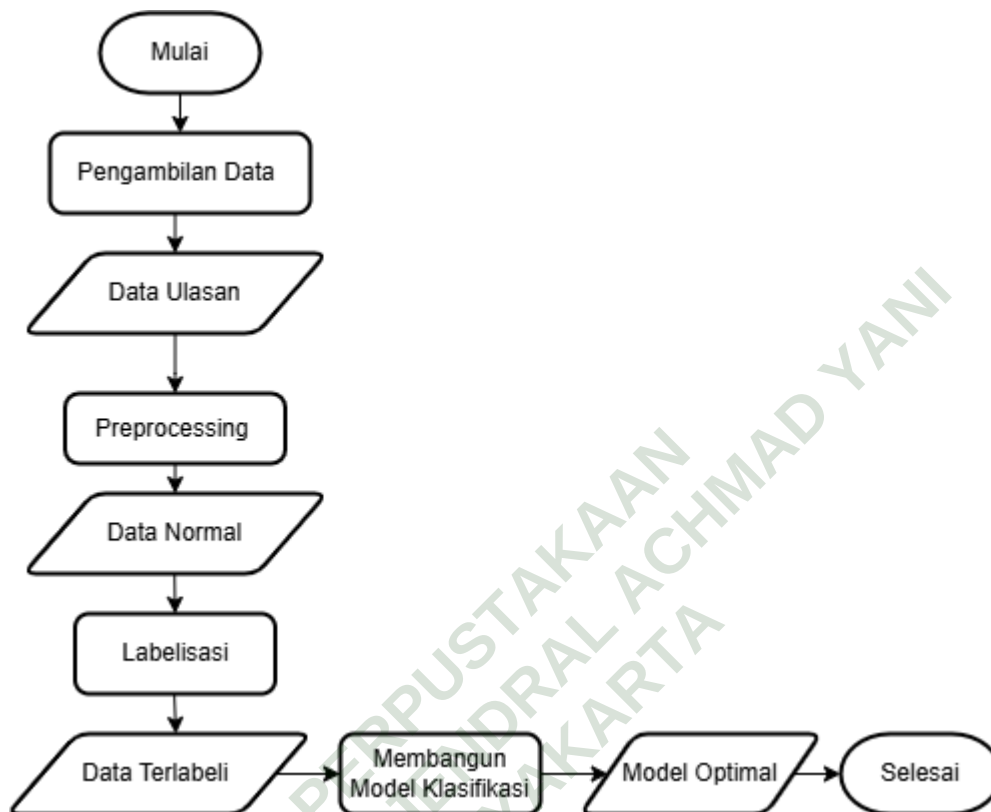
Pengumpulan data dilakukan berfungsi untuk mengambil data rating dan komentar pengguna aplikasi Vidio dan Loklok yang tersedia di Google Play Store. Proses ini melibatkan teknik web scraping dengan bantuan Google Collab, dimana data yang dikumpulkan berupa komentar dalam bahasa Indonesia.

Sistem perangkat lunak dan berbagai aplikasi yang diterapkan di pembuatan sistem meliputi:

1. Sistem Operasi: Windows 11.
2. Google Collab: Untuk mengambil data rating dan ulasan di PlayStore.
3. Bahasa Pemrograman Python 3.12.7
4. Jupyter Notebook 7.2.2
5. Microsoft Office Excel 2013

3.2 JALAN PENELITIAN

Penelitian ini melalui beberapa tahap yang dapat dilihat pada Gambar 3.1



Gambar 3. 1 Flowchart Jalan Penelitian

Penjelasan tentang jalannya Penelitian ini adalah:

1. Pengambilan data di Google Play Store

Proses pengumpulan data diperoleh dari rating dan ulasan aplikasi Vidio dan Loklok pada Google Play Store dengan teknik *scraping* data menggunakan Google Collab. *Scraping* digunakan untuk mengekstrak data secara otomatis. Data diambil melalui *scraping* dapat berupa teks dan informasi lainnya dan menghasilkan format dalam bentuk CSV (*Comma-Separated Values*).

2. Data Ulasan

Data ulasan merupakan informasi atau komentar yang diberikan oleh pengguna mengenai produk, layanan atau aplikasi. Data ini berupa teks yang menggambarkan pengalaman, opini, atau kepuasan pengguna terhadap hal yang dialami.

3. Preprocessing

Tahap *Preprocessing* sangat penting dalam pemrosesan data yang bertujuan untuk mengatasi berbagai masalah yang dapat memengaruhi hasil analisis. Proses ini mencakup beberapa langkah yang harus diterapkan guna memastikan data yang digunakan lebih bersih dan akurat, antara lain sebagai berikut:

a. *Cleaning*

Pada tahap ini proses yang dilakukan *cleaning*, untuk menghapus karakter non alfabet dalam suatu teks sehingga teks tersebut menjadi lebih mudah diolah. Seperti menghapus *emoticon*, angka, *tagar*, mention dan menghapus URL yang terdapat pada komentar.

b. *Filtering*

Untuk menghapus tanda baca yang diterapkan kata kunci serta menghapus beberapa kata yang tidak perlu.

c. *Stemming*

Tahap *stemming* dilakukan untuk mengonversi kalimat berimbuhan dengan mengubahnya ke dalam bentuk dasar.

d. *Tokenizing*

Langkah *tokenizing* digunakan untuk membagi teks ke dalam bagian-bagian kecil, seperti kata atau frasa, sehingga dapat lebih mudah diproses dan dianalisis.

4. Data Normal

Proses normalisasi data yang bertujuan untuk menyesuaikan format teks agar konsisten dan mudah diproses. Proses ini berupa penggantian simbol atau emotikon, perbaikan ejaan dan lainnya.

5. Labelisasi

Proses pemberian kategori atau label pada data untuk mengklasifikasikannya dalam kelompok tertentu. Berikut ini label yang digunakan dalam penelitian ini adalah:

a. Label Positif

Digunakan untuk ulasan yang menyatakan kepuasan baik atau pengalaman baik dari pengguna.

b. Label Negatif

Digunakan untuk ulasan yang mengungkapkan ketidakpuasan atau pengalaman buruk dari pengguna.

6. Data Terlabeli

Proses data yang telah terkumpul akan diberikan label sesuai kategori yang dibutuhkan dalam penelitian, misalnya sentimen “positif”, dan “negatif”.

7. Membangun Model Klasifikasi

Model klasifikasi dibangun menggunakan algoritma Naive Bayes untuk mengklasifikasi sentimen ulasan pengguna terhadap aplikasi Vidio dan Loklok ke dalam kategori positif dan negatif. Proses ini dimulai dengan data yang telah melalui tahapan preprocessing dan labelisasi.

a. Pelatihan (*Training*)

Pada tahap ini, data yang sudah diberikan label digunakan untuk melatih model serta mengidentifikasi pola-pola dalam teks ulasan dan hubungan dengan label sentimen yang diberi.

b. Pengujian (*Testing*)

Pada tahap ini, data yang telah dilatih diuji menggunakan data yang tidak pernah digunakan pada tahap pelatihan untuk mengukur akurasi dan kinerja model. Evaluasi dilakukan dengan menggunakan metrik seperti akurasi, presisi, recall, dan f1-score berguna untuk menilai sejauh mana model dapat mengklasifikasikan data secara tepat dan konsisten.

8. Model Optimal

Setelah model dibangun, dilakukan evaluasi menggunakan confusion metrik seperti akurasi, presisi, recall, f1-score untuk mencari parameter dan menghasilkan suatu model yang optimal.