

BAB III

METODE PENELITIAN

Metodologi penelitian merupakan tahapan yang dilakukan selama pelaksanaan penelitian untuk memastikan proses berjalan dengan baik dan terstruktur, serta mencapai tujuan penelitian yang ditetapkan. Berikut adalah alur dari metodologi penelitian ini.



Gambar 3. 1 Tahapan Penelitian

1.1 Pengumpulan Data

Pengumpulan dataset dilakukan menggunakan metode *web scraping* dengan bahasa pemrograman *Python* dengan pustaka *google-play-scraper*. Data yang diambil adalah ulasan yang berasal dari wilayah Indonesia dan ditulis dalam bahasa Indonesia bersumber dari *play store*.

1.2 Preprocessing Data

Text preprocessing merupakan proses yang bertujuan untuk mengolah data teks agar menjadi lebih terstruktur, melalui serangkaian tahapan yang mencakup case folding, tokenisasi, penyaringan, dan stemming. Namun, tidak ada pedoman baku mengenai setiap langkah dalam text preprocessing, karena proses ini merupakan salah satu bentuk implementasi dari text mining.

1.3 TF-IDF Dan Pembagian Data

1.3.1 TF-IDF

Pada tahap ini, dilakukan untuk memperoleh nilai atau bobot dari kata-kata yang telah diekstraksi. Metode ini bergantung pada kemunculan kata (term) dalam dokumen dengan mengalikan frekuensi term (TF) dengan invers dari frekuensi dokumen (IDF). Frekuensi kemunculan sebuah term akan mempengaruhi penilaian bobotnya.

1.3.2 Pembagian Data

Sebelum melakukan pemodelan, data akan dibagi dua bagian yaitu data latih (*training*) 80% dan data uji (*testing*) 20%. Proses pembagian data tersebut dapat dilakukan menggunakan salah satu teknik sederhana yaitu hold-out validation. Presentase tersebut terbagi 80:20. Data pelatihan digunakan untuk melatih model agar dapat mengidentifikasi pola dalam data, sementara data pengujian digunakan untuk mengevaluasi dan menilai sejauh mana kemampuan model yang telah dilatih.

1.4 Support Vector Machine (SVM)

Dalam hal ini, untuk menentukan bobot atau nilai suatu kata, dapat dilakukan dengan mengikuti jenis TF berdasarkan frekuensi kemunculan kata term dalam dokumen dengan memberikan nilai frekuensi term (TF) dikalikan dengan kebalikannya, yaitu frekuensi dokumen (IDF). Frekuensi kemunculan suatu term akan menjadi dasar penentuan nilai bobotnya.

1.5 Evaluasi Model Dan Analisis

1.5.1 Evaluasi Model

Model di evaluasi berdasarkan klasifikasi melalui matriks kebingungan sehingga akurasi, presisi, dan perolehan kembali model dapat diukur secara kuantitatif dalam hal klasifikasi. Rumus berikut digunakan.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \dots\dots\dots(1)$$

$$\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP}) \dots\dots\dots(2)$$

$$\text{Recall} = (\text{TP}) / (\text{TP} + \text{FN}) \dots\dots\dots(3)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \dots\dots\dots(4)$$

1.5.2 Analisis

Analisis ini bertujuan untuk mengidentifikasi dan memahami seberapa sering kata-kata tertentu muncul dalam setiap kategori sentimen pada ulasan aplikasi pinjaman online. Proses ini di visualisasikan menggunakan teknik wordcloud, sehingga memudahkan dalam memahami opini pengguna secara menyeluruh tanpa harus membaca setiap komentar satu per satu.