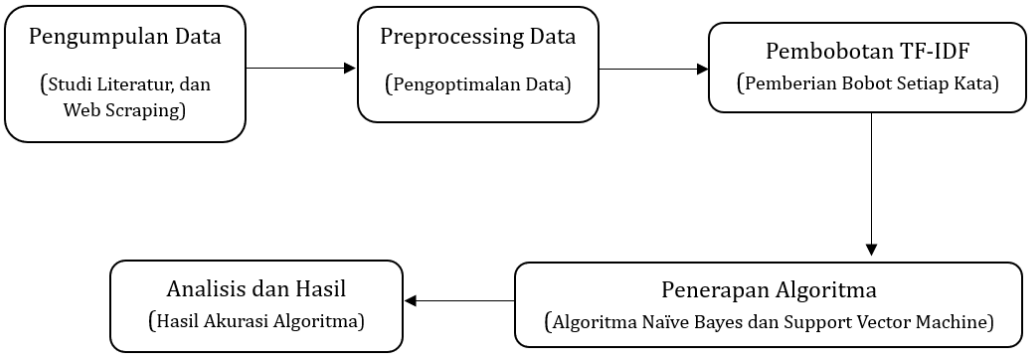
BAB III
METODE PENELITIAN

Penelitian ini dilakukan pada aplikasi Grab Indonesia dengan menggunakan metode *Naïve Bayes* dan *Support Vector Machine* (SVM) untuk analisis data. Data dikumpulkan melalui proses scraping ulasan pengguna dari Google *Play Store*. Selanjutnya, data tersebut diolah untuk analisis sentimen terhadap ulasan pengguna aplikasi Grab di Google *Play Store* menggunakan bahasa *Python* pada platform *Google Colab*.

3.1 Kerangka Penelitian



Gambar 1. Alur Penelitian

Penelitian diawali dengan pengumpulan data ulasan pengguna aplikasi Grab Indonesia dari Google *Play Store*, yang kemudian melalui tahap *preprocessing* data seperti pembersihan teks, tokenisasi, dan penghapusan kata-kata yang tidak relevan. Data yang telah diolah selanjutnya diberi label sentimen (positif atau negatif) untuk dijadikan dataset analisis. Metode utama yang digunakan adalah algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM), yang dibandingkan untuk mengevaluasi performa masing-masing dalam klasifikasi sentimen, menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*.

3.2 Pengumpulan Data

Dalam pengembangan penelitian ini, penulis mengumpulkan data dan informasi yang mendukung proses pengumpulan data yang relevan, sebagaimana dijelaskan berikut:

a. Studi Literatur

Penulis melakukan studi literatur dengan mengumpulkan dan menganalisis berbagai sumber relevan, seperti artikel dan jurnal ilmiah, yang berkaitan dengan topik penelitian. Fokus utama studi literatur mencakup teori analisis sentimen, penerapan metode *Naïve Bayes Classifier* dan *Support Vector Machine* (SVM), serta penggunaan perangkat lunak seperti *Google Colab* dalam proses analisis. Selain itu, penulis juga mengacu pada jurnal-jurnal terpercaya dan situs web resmi sebagai sumber informasi tambahan yang mendukung keakuratan dan kelengkapan data. Langkah ini dilakukan untuk membangun dasar teoritis yang kuat, memperdalam pemahaman, dan memastikan penelitian dilakukan secara sistematis dan terarah.

b. Web Scraping

Penulis mengumpulkan data ulasan pengguna aplikasi Grab Indonesia yang tersedia di Google *Play store* melalui proses *web scraping*. Untuk mendukung proses tersebut digunakan alat bantu berupa perangkat lunak atau ekstensi *Google Chrome* bernama *Instant Data Scraper*. Proses pengumpulan data ini berhasil menghimpun sebanyak 2.140 ulasan pengguna yang kemudian disimpan dalam format file *.xlsx* untuk mempermudah pengolahan dan analisis lebih lanjut. Langkah ini dilakukan secara sistematis untuk memastikan data yang diperoleh mencerminkan pandangan pengguna secara representatif.

3.3 Proses Implementasi

Pada proses implementasi, penulis menggambarkan penerapan analisis sentimen dengan menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine* (SVM), yang diawali dengan serangkaian langkah preprocessing data.

a. Dataset

Penelitian ini menggunakan ulasan pengguna aplikasi Grab Indonesia di Google *Playstore* sebagai sumber data utama. Data tersebut diperoleh melalui teknik *web scraping*, dengan bantuan ekstensi *Google Chrome* bernama *Instant Data Scraper*. Setelah proses pengambilan data selesai, dataset yang berhasil dikumpulkan disimpan dalam format *.xlsx* untuk memudahkan pengelolaan lebih lanjut. Sebelum data memasuki tahap

preprocessing, dilakukan pemisahan *dataset* menjadi data pelatihan (training data) dan data pengujian (testing data) guna mendukung analisis sentimen yang lebih terstruktur dan akurat. Tahapan ini dirancang untuk memastikan kualitas data yang digunakan dalam proses analisis.

b. Pre-Processing

Tahap *preprocessing* bertujuan untuk menghilangkan elemen-elemen yang tidak relevan dalam *dataset*, sehingga data menjadi lebih bersih dan siap untuk dianalisis (Syahril Dwi Prasetyo et al., 2023). Proses ini mirip dengan tahap pengolahan data, namun lebih spesifik dalam penanganan teks. Dalam penelitian ini, *preprocessing* mencakup beberapa langkah penting, yaitu *tokenizing*, *transform cases*, *stemming*, dan *filtering*. *Tokenizing* berfungsi memecah teks menjadi unit-unit kata, sementara *transform cases* mengubah semua huruf menjadi format huruf kecil untuk konsistensi. *Stemming* dilakukan untuk mengembalikan kata ke bentuk dasarnya, dan *filtering* digunakan untuk menghapus kata-kata yang tidak bermakna atau tidak relevan. Semua langkah ini diterapkan dengan menggunakan perangkat lunak *Google Colab* untuk memastikan proses dilakukan secara efisien dan akurat.

c. Split Data

Langkah selanjutnya dalam proses penelitian adalah *split data*, yang bertujuan untuk membagi dataset yang telah diproses menjadi dua bagian yang terpisah, yaitu data latih (training data) dan data uji (testing data). Pembagian ini sangat penting untuk memastikan bahwa model yang dibangun dapat dievaluasi secara objektif. Data latih digunakan untuk melatih model, sementara data uji digunakan untuk menguji kinerja model yang telah dilatih, dengan tujuan untuk menilai seberapa baik model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya.

d. Pelabelan Sentimen

Proses pelabelan sentimen dilakukan dengan membagi *dataset* yang telah diproses menjadi dua bagian, yaitu data latih dan data uji, dengan komposisi pembagian sebesar 80% untuk data latih dan 20% untuk data uji. Data latih digunakan untuk melatih model dalam mengidentifikasi sentimen

dari ulasan pengguna, sedangkan data uji berfungsi untuk menguji kemampuan model dalam mengklasifikasikan sentimen pada data yang belum pernah dilihat sebelumnya.

e. Pengujian Data

Setelah tahap *preprocessing* selesai, langkah berikutnya adalah pengujian data melalui klasifikasi menggunakan dua metode utama, yaitu *Naïve Bayes Classifier* dan *Support Vector Machine* (SVM). Proses klasifikasi ini dilakukan dengan memanfaatkan perangkat lunak *Google Colab* untuk menjalankan analisis secara efisien. Hasil dari proses klasifikasi ini akan digunakan dalam perhitungan *confusion matrix*, yang merupakan tahap penting untuk mengukur berbagai metrik evaluasi model, seperti akurasi, presisi, dan *recall*. Pengukuran ini bertujuan untuk mengevaluasi kinerja kedua metode klasifikasi yang diuji dan memberikan gambaran tentang seberapa efektif model dalam mengidentifikasi sentimen yang benar, serta sejauh mana model mampu meminimalkan kesalahan dalam pengklasifikasian data.

f. Analisis Hasil

Hasil dari analisis sentimen ini akan dievaluasi secara mendalam untuk mempermudah pemahaman terhadap data yang telah diperoleh. Evaluasi ini akan mencakup perbandingan komprehensif antara kedua metode yang digunakan, yaitu *Naïve Bayes Classifier* dan *Support Vector Machine* (SVM), untuk menilai kelebihan dan kekurangan masing-masing dalam mengklasifikasikan sentimen. Perbandingan ini tidak hanya akan memberikan wawasan tentang kinerja masing-masing metode, tetapi juga akan mengungkapkan metode mana yang lebih efektif dalam konteks analisis ulasan aplikasi Grab Indonesia. Hasil evaluasi ini akan menjadi dasar untuk menyimpulkan temuan akhir penelitian, yang kemudian digunakan untuk memberikan rekomendasi dan pemahaman yang lebih baik mengenai sentimen pengguna terhadap aplikasi tersebut.