

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian ini bertujuan untuk menganalisis sentimen dari ulasan aplikasi Kredivo yang diambil dari Google Play Store. Analisis dilakukan dengan menggunakan algoritma Support Vector Machine (SVM), Naïve Bayes, dan Random Forest untuk mengklasifikasikan ulasan pengguna ke dalam dua kategori utama, yaitu sentimen positif dan negatif. Diharapkan hasil penelitian ini dapat memberikan insight bagi pengembang Kredivo dalam memahami pandangan pengguna terhadap layanan mereka, sehingga menjadi dasar untuk meningkatkan kualitas layanan dan pengalaman pengguna.

Alur penelitian ini terdiri dari beberapa tahap utama, yaitu identifikasi masalah, perumusan masalah, pengambilan data, penentuan tujuan penelitian, preprocessing data, pembagian data untuk pelatihan dan pengujian, pengujian algoritma SVM, Naïve Bayes, dan Random Forest, proses pemodelan, evaluasi hasil algoritma, dan *deployment*. Setiap tahap dirancang untuk memastikan proses penelitian dilakukan secara terstruktur dan sistematis.

Ulasan aplikasi Kredivo dipilih sebagai objek penelitian karena aplikasi ini merupakan platform pinjaman online yang populer di Indonesia dengan total ulasan yang sangat besar yang mencapai 4,13 juta ulasan. Ulasan ini mencerminkan pengalaman pengguna terhadap proses pengajuan, transparansi biaya, suku bunga, dan layanan pelanggan. Dengan dataset ulasan yang besar dan beragam, analisis ini memungkinkan pengujian performa algoritma SVM, Naïve Bayes, dan Random Forest pada data teks berdimensi tinggi, sekaligus memberikan hasil yang relevan untuk konteks pengembangan layanan keuangan berbasis aplikasi.

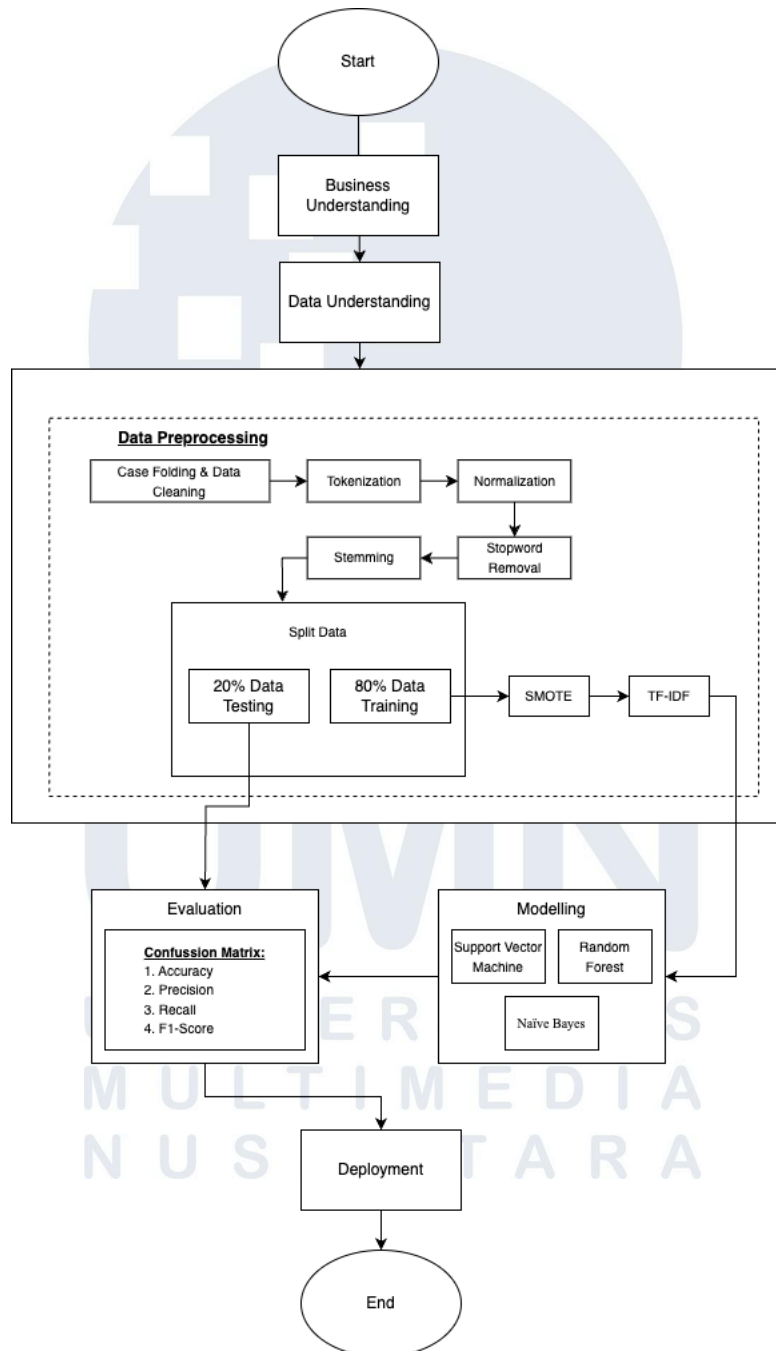
Penelitian ini juga diharapkan dapat memberikan kontribusi terhadap kemajuan teknologi analisis sentimen khususnya di bidang fintech., khususnya

dalam memahami kepuasan pengguna dan meningkatkan kualitas layanan berbasis ulasan konsumen.

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Alur Penelitian pada aplikasi Kredivo akan diawali dengan kerangka pikir yang telah disesuaikan dengan Framework CRISP-DM:



Gambar 3. 1 Alur Penelitian

3.2.1.1 Business Understanding

Tahap Business Understanding bertujuan untuk memahami dan merumuskan tujuan penelitian berdasarkan permasalahan yang dihadapi oleh pengguna dalam memberikan ulasan pada aplikasi Kredivo di platform Google Play Store. Kredivo adalah aplikasi layanan kredit digital yang banyak digunakan oleh masyarakat Indonesia untuk melakukan transaksi pembelian secara cicilan tanpa kartu kredit. Dalam beberapa tahun terakhir, popularitas aplikasi ini meningkat pesat seiring dengan kebutuhan masyarakat akan layanan keuangan yang cepat dan mudah diakses.

Dalam proses penggunaan aplikasi, pengguna sering kali memberikan ulasan di Google Play Store yang berisi pendapat, pengalaman, dan keluhan mereka terhadap layanan yang disediakan oleh Kredivo. Ulasan ini mencakup aspek positif, , positif dan negatif terkait dengan kemudahan pengajuan kredit, kecepatan persetujuan, bunga pinjaman, layanan pelanggan, hingga keamanan data. Ulasan ini menjadi sumber informasi penting yang dapat digunakan untuk meningkatkan kualitas layanan dan memperbaiki kelemahan yang ada pada aplikasi.

Penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi Kredivo dengan memanfaatkan tiga algoritma machine learning, yakni Support Vector Machine (SVM), Naïve Bayes, dan Random Forest. Tujuan utama dari penelitian ini adalah:

1. Mengelompokkan sentimen ulasan ke dalam dua kategori utama yaitu, positif dan negatif.
2. Membandingkan performa dari ketiga algoritma yang digunakan dalam proses klasifikasi sentimen.

3. Memberikan rekomendasi algoritma terbaik berdasarkan hasil evaluasi performa dalam hal akurasi, *precision*, *recall*, dan *F1-Score*.

Penelitian ini diharapkan memberikan kontribusi penting bagi pengembang Kredivo dalam meningkatkan kualitas layanan berdasarkan hasil analisis sentimen pengguna. Selain itu, penelitian ini juga diharapkan dapat memperkaya pengembangan metode analisis sentimen berbasis machine learning yang dapat diaplikasikan pada platform ulasan lain di Google Play Store.

3.2.1.2 Data Understanding

Tahap *Data Understanding* bertujuan untuk mengenali dan memahami karakteristik dataset yang digunakan dalam penelitian ini, yang berasal dari ulasan pengguna aplikasi Kredivo di Google Play Store. Dataset ini berisi total 10.000 ulasan, yang diambil dari Google Play Store. Dalam tahap ini, dilakukan eksplorasi awal untuk memahami struktur data, pola distribusi rating, dan potensi permasalahan yang mungkin muncul dalam proses analisis sentimen. Langkah-langkah eksplorasi data meliputi:

a) DistribusiRatingUlasan:

Ulasan diklasifikasikan berdasarkan rating yang diberikan pengguna di Google Play Store, yaitu dari skala 1 hingga 5 [10]. Distribusi rating ini akan digunakan sebagai dasar untuk mengkategorikan sentimen, di mana:

- a. Rating 4–5 dikategorikan sebagai sentimen positif.
- b. Rating 1–2 dikategorikan sebagai sentimen negatif.

b) PembersihanAwalData:

Dilakukan identifikasi dan penghapusan simbol, karakter tidak valid, serta tanda baca yang tidak relevan untuk meningkatkan kualitas teks ulasan. Ulasan yang sangat

pendek atau tidak bermakna juga dihapus agar tidak memengaruhi hasil analisis.

Eksplorasi data ini bertujuan untuk memahami pola dasar dari dataset, memastikan kualitas data, dan menyiapkan data untuk tahap selanjutnya. Hasil dari tahap ini akan digunakan dalam proses *pre-processing* dan pemodelan untuk meningkatkan akurasi dalam mengklasifikasikan sentimen berdasarkan rating yang diberikan pengguna.

3.2.1.3 Data Preparation

Tahap Data Preparation bertujuan untuk mempersiapkan data ulasan dari aplikasi Kredivo agar siap digunakan dalam proses pemodelan analisis sentimen. Data yang telah dikumpulkan dari Google Play Store melewati serangkaian proses untuk memastikan kualitas data, mengurangi noise, dan meningkatkan akurasi model. Tahap Data Preparation dalam penelitian ini meliputi proses Data Preprocessing, pembagian data (*data splitting*), serta konversi data menjadi bentuk numerik menggunakan teknik TF-IDF. Berikut adalah langkah-langkah yang dilakukan dalam tahap Data Preparation:

1) Case folding & Data cleaning

Pada tahap ini, seluruh teks ulasan diubah menjadi huruf kecil (lowercase) guna menghilangkan perbedaan yang timbul akibat penggunaan huruf kapital. Selain itu, simbol, angka, tanda baca, dan karakter tidak relevan dihapus menggunakan teknik Regular Expressions (Regex) untuk membersihkan teks dari karakter yang tidak relevan dengan menerapkan pola Regex `[^a-zA-Z ]`. Pola ini bekerja dengan cara menghapus semua karakter kecuali huruf (a-z, A-Z) dan spasi, sehingga secara efektif menghilangkan angka, tanda baca, dan simbol seperti titik (.), koma (,), tanda seru (!), tanda tanya (?), tagar (#), simbol at (@), hingga berbagai macam emoji. Teks yang mengandung URL dan

alamat email juga dihapus untuk menghindari gangguan pada hasil analisis.

2) *Tokenization*

Setelah proses pembersihan data, teks ulasan dipecah menjadi token atau kata-kata individu memanfaatkan pustaka Natural Language Toolkit (NLTK). Proses ini bertujuan untuk memisahkan setiap kata dalam ulasan sehingga dapat dianalisis secara terpisah oleh model.

3) *Normalization*

Pada tahap normalisasi, kata-kata yang memiliki makna sama tetapi memiliki perbedaan dalam penulisan diseragamkan. Sebagai contoh, kata "gak", "nggak", dan "tidak" akan disamakan menjadi "tidak" untuk meningkatkan konsistensi data.

4) *Stopword removal*

Kata-kata umum yang tidak memiliki makna penting seperti "dan", "atau", "dengan", dan lainnya dihapus menggunakan daftar stopwords bahasa Indonesia dari library NLTK. Penghapusan stopwords bertujuan untuk mengurangi dimensi data dan memusatkan perhatian pada kata-kata yang memiliki arti penting dalam analisis sentimen.

5) *Stemming*

Pada tahap ini, kata-kata dalam ulasan diubah ke bentuk dasarnya dengan menggunakan pustaka Sastrawi. Misalnya, kata "bermain", "bermainlah", dan "bermainnya" akan dikonversi menjadi "main". Proses ini bertujuan untuk menyederhanakan analisis dan meningkatkan kesamaan data.

6) *Split Data*

Setelah data selesai diproses, data kemudian dibagi menjadi dua bagian, yaitu:

- a) Data Training, sebanyak 80% dari total data, digunakan untuk melatih model.

- b) Data Testing: sebanyak 20% dari total data, digunakan untuk menguji kinerja model.

Pembagian data ini dilakukan secara acak (random split) agar distribusi data pada masing-masing kelompok menjadi merata.

7) **SMOTE**

Setelah data dibagi menjadi data latih dan data uji, langkah selanjutnya adalah menangani masalah ketidakseimbangan kelas yang teridentifikasi pada data latih. Untuk mengatasi hal ini, akan diterapkan teknik *oversampling* SMOTE. Penting untuk dicatat bahwa SMOTE hanya akan diaplikasikan pada data latih (80%) untuk menyeimbangkan distribusi kelas sebelum data dimasukkan ke dalam model. Langkah ini bertujuan untuk mencegah kebocoran data (*data leakage*) dan memastikan data uji (20%) tetap dapat digunakan sebagai representasi data asli yang murni untuk evaluasi.

8) **TF-IDF (Term Frequency-Inverse Document Frequency)**

Teks yang sudah diproses kemudian dikonversi menjadi bentuk numerik dengan menggunakan metode *TF-IDF*. Teknik ini memberikan bobot lebih besar pada kata-kata yang sering muncul dalam satu dokumen, namun jarang ditemukan di seluruh dataset. Teknik ini bertujuan untuk meningkatkan kemampuan model dalam membedakan makna kata berdasarkan konteks penggunaannya.

Hasil dari tahap Data Preparation ini adalah dataset teks yang bersih, seragam, dan telah direpresentasikan dalam bentuk numerik. Dataset ini siap digunakan dalam tahap pemodelan untuk membangun model analisis sentimen berbasis algoritma Support Vector Machine (SVM), Naïve Bayes, dan Random Forest. Langkah ini memastikan bahwa model dapat belajar dari data yang relevan dan berkualitas, sehingga meningkatkan performa analisis sentimen secara keseluruhan.

3.2.1.4 **Modelling**

Pada tahap ini, dilakukan analisis pemodelan menggunakan algoritma klasifikasi yang dipilih, yaitu Naive Bayes, Support Vector Machines (SVM), dan Random Forest. Pemilihan algoritma tersebut didasarkan pada penelitian sebelumnya yang telah berhasil menerapkan ketiga metode ini dalam analisis sentiment [6]. Proses pemodelan akan diimplementasikan secara efektif menggunakan bahasa pemrograman Python.

3.2.1.5 Evaluation

Pada tahap ini, dilakukan evaluasi kinerja model yang telah dikembangkan. Evaluasi meliputi penyajian metrik akurasi, *precision*, *recall*, *F1-Score*, serta *confusion matrix* yang diperoleh dari algoritma Naive Bayes, Support Vector Machines (SVM), dan Random Forest, termasuk hasil pengujian model.

3.2.1.6 Deployment

Tahap terakhir dalam proses ini adalah penerapan hasil yang diperoleh. Pada penelitian ini, implementasi dilakukan dengan membuat sebuah dashboard yang menampilkan visualisasi data analisis sentimen terhadap opini pengguna mengenai aplikasi Kredivo. Dashboard tersebut bertujuan memberikan gambaran yang jelas dan mudah dipahami tentang ulasan masyarakat terkait aplikasi tersebut.

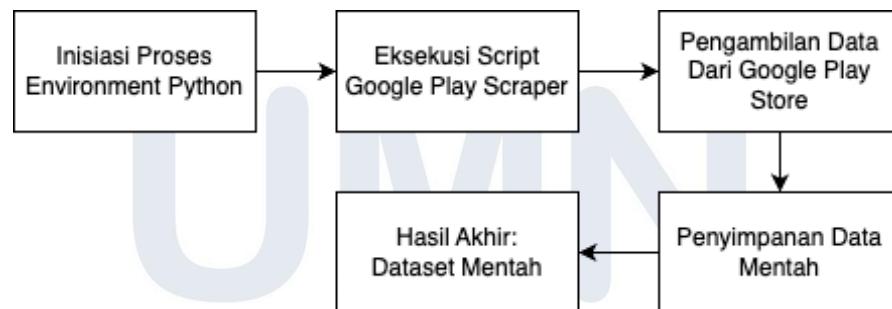
3.3 Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python yang dijalankan di Jupyter Notebook dengan dukungan Anaconda Navigator sebagai lingkungan kerja. Data yang dikumpulkan berupa ulasan pengguna aplikasi Kredivo yang diambil dari Google Play Store menggunakan pustaka *google-play-scraper*. Proses pengambilan data dilakukan secara otomatis melalui pemrograman, sehingga memudahkan dalam memperoleh data dalam jumlah besar secara efisien dan sistematis. Dalam proses ini berhasil dikumpulkan sebanyak 10.000 data ulasan, dan data tersebut disesuaikan dengan periode waktu penelitian yaitu pada tahun 2025. Pemilihan periode ini bertujuan untuk menangkap opini

pengguna yang paling mutakhir dan relevan dengan kondisi serta fitur aplikasi Kredivo terkini, sehingga hasil analisis sentimen dapat memberikan wawasan yang *up-to-date* bagi pengembang dan pihak terkait. Data yang diperoleh mencakup rating, isi ulasan, dan tanggal ulasan, yang kemudian disimpan dalam format CSV untuk memudahkan pengolahan lebih lanjut pada tahap preprocessing dan modeling.

3.3.1 Bagan Pengumpulan Data

Proses pengumpulan hingga persiapan data untuk pemodelan dalam penelitian ini mengikuti serangkaian tahapan yang terstruktur, yang diilustrasikan secara visual pada Gambar 3.2. Proses ini diawali dengan tahap persiapan alat (*Mempersiapkan Tools*), di mana lingkungan kerja dan pustaka yang dibutuhkan disiapkan, meliputi bahasa pemrograman Python, Jupyter Notebook, dan pustaka google-play-scraper. Selanjutnya, dilakukan proses *scraping* untuk mengumpulkan 10.000 ulasan dari Google Play Store, yang kemudian hasilnya disimpan sebagai dataset ulasan mentah dalam format file CSV.



Dataset mentah yang telah dikumpulkan belum dapat langsung digunakan untuk pemodelan karena masih mengandung *noise* dan format teks yang tidak seragam, sehingga harus melalui tahap *Preprocessing*. Tahap krusial ini melibatkan serangkaian proses pembersihan dan standardisasi teks yang terdiri dari *Case Folding & Data Cleaning*, *Tokenization*, *Normalization*, *Stopword Removal*, dan *Stemming*. Setelah melalui seluruh tahapan tersebut, hasil akhirnya

adalah sebuah *Dataset* Bersih yang siap digunakan untuk tahap selanjutnya, yaitu ekstraksi fitur dan pemodelan *machine learning*.

3.4 Teknik Analisis Data

Penelitian ini menggunakan pendekatan kuantitatif untuk menganalisis data dengan tujuan mengukur dan mengevaluasi sentimen pengguna secara sistematis berdasarkan ulasan dari Google Play Store. Pemrosesan dan analisis data dilakukan dengan bahasa pemrograman Python melalui beberapa tahapan, mulai dari preprocessing teks, ekstraksi fitur menggunakan TF-IDF, hingga pembangunan model klasifikasi. Tiga algoritma machine learning Support Vector Machine (SVM), Naïve Bayes, dan Random Forest—diimplementasikan dan dibandingkan untuk mendapatkan hasil klasifikasi sentimen yang paling akurat. Setiap algoritma dievaluasi menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-Score* agar dapat menentukan algoritma yang paling efektif dalam mengklasifikasikan sentimen ulasan aplikasi Kredivo.

