

BAB 3

METODOLOGI PENELITIAN

Dalam penelitian ini menggunakan metode klasifikasi algoritma Multinomial Naïve Bayes. Proses dalam implementasi sistem ini akan menggunakan istilah frekuensi untuk mencari pola algoritma dengan melakukan alur perhitungan algoritma Multinomial Naïve Bayes[Huda, Setiawan, Manongga, Purnomo, Hendry & Tukino 2024]. Pendekatan kuantitatif digunakan untuk mengukur kinerja algoritma secara objektif melalui evaluasi metrik seperti akurasi, presisi, recall, dan F1-score. Tujuan utama penelitian ini adalah membangun sebuah sistem yang mampu melakukan pengambilan data (scraping), pra-pemrosesan teks, ekstraksi fitur, pelatihan model, serta evaluasi hasil klasifikasi terhadap ulasan pengguna aplikasi Jaket Boat.

3.1 Objek Penelitian

Penelitian ini dilakukan dengan objek utama berupa ulasan pengguna aplikasi Jaket Boat yang tersedia di platform Google Play Store. Jaket Boat merupakan aplikasi layanan pemesanan tiket kapal secara daring yang mempermudah masyarakat dalam melakukan perjalanan laut, terutama di wilayah kepulauan seperti Kepulauan Seribu. Dengan semakin banyaknya pengguna yang mengunduh dan menggunakan aplikasi tersebut, muncul pula ribuan ulasan yang mencerminkan pengalaman pengguna terhadap layanan yang diberikan. Data ulasan tersebut menjadi sangat berharga karena berisi persepsi, keluhan, dan kepuasan pengguna secara langsung. Objek penelitian ini dipilih karena data yang tersedia bersifat publik, aktual, dan merepresentasikan opini pengguna secara langsung tanpa intervensi. Penelitian ini memanfaatkan data ulasan sebagai sumber utama untuk dianalisis menggunakan pendekatan analisis sentimen berbasis machine learning, khususnya algoritma Naïve Bayes, dengan tujuan untuk mengklasifikasikan sentimen ulasan menjadi kategori positif dan negatif secara otomatis. Adapun batasan dalam penelitian ini antara lain:

1. Data yang digunakan hanya berupa ulasan dalam bahasa Indonesia dari pengguna aplikasi Jaket Boat di Google Play Store;

2. Sentimen yang diklasifikasikan dibatasi hanya pada dua kelas yaitu positif dan negatif tanpa mempertimbangkan sentimen netral;
3. Algoritma yang digunakan hanya algoritma Multinomial Naïve Bayes, tanpa perbandingan dengan algoritma lain;
4. Proses scraping dilakukan dalam periode waktu tertentu, sehingga data bersifat snapshot dan tidak mewakili seluruh populasi pengguna aplikasi.

3.1.1 Deskripsi Aplikasi Jaket Boat

Aplikasi Jaket Boat merupakan layanan transportasi laut berbasis digital yang dikelola untuk melayani pemesanan tiket kapal secara daring. Layanan ini menyediakan aksesibilitas yang lebih mudah dan praktis bagi masyarakat, khususnya yang berada di wilayah perairan seperti Kepulauan Seribu. Aplikasi ini dapat diunduh melalui Google Play Store, di mana pengguna dapat meninggalkan ulasan terkait pengalaman mereka menggunakan aplikasi tersebut. Ulasan ini menjadi cerminan langsung dari performa aplikasi dan kepuasan pengguna, yang kemudian digunakan dalam penelitian ini untuk mengembangkan model klasifikasi sentimen secara otomatis.

Gambar 3. 1 Logo Aplikasi Jaket Boat



3.1.2 Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan secara daring (online) dengan mengakses dan memproses data yang tersedia di internet. Proses pengumpulan data dilakukan dari platform Google Play Store, sedangkan analisis data dilakukan menggunakan lingkungan pemrograman berbasis cloud yaitu Google Colaboratory yang mendukung bahasa Python. Penelitian ini dilakukan selama semester genap tahun

akademik 2024/2025, mulai dari tahap pengumpulan data, pengolahan, pelabelan, analisis, hingga evaluasi model.

3.2 Tahapan Penelitian

Prosedur penelitian ini terdiri atas beberapa tahapan yang saling berkaitan mulai dari pengumpulan data hingga evaluasi hasil. Prosedur dimulai dengan proses pengumpulan data ulasan pengguna aplikasi Jaket Boat dari Google Play Store melalui teknik scraping. Selanjutnya, data yang diperoleh akan diberi label sentimen berdasarkan skor rating yang diberikan pengguna, lalu dilakukan tahapan preprocessing teks agar data lebih siap digunakan dalam proses pembelajaran mesin. Setelah data bersih, dilakukan proses ekstraksi fitur menggunakan metode TF-IDF untuk mengubah data teks menjadi bentuk numerik. Langkah selanjutnya adalah penerapan algoritma Naïve Bayes untuk melakukan klasifikasi sentimen, dan diakhiri dengan proses evaluasi model menggunakan metrik akurasi, precision, recall, dan F1-score.

3.2.1 Teknik Pengumpulan Data

Tahap awal penelitian adalah pengumpulan data berupa ulasan pengguna dari aplikasi Jaket Boat. Data dikumpulkan melalui proses web scraping menggunakan bahasa pemrograman Python dan pustaka seperti requests dan BeautifulSoup. Ulasan dikumpulkan dari halaman aplikasi Jaket Boat yang tersedia di platform Google Play Store (<https://www.kapaltiket.com>), khususnya review mengenai permasalahan aplikasi saat ini.

Langkah-langkah pada tahap ini meliputi :

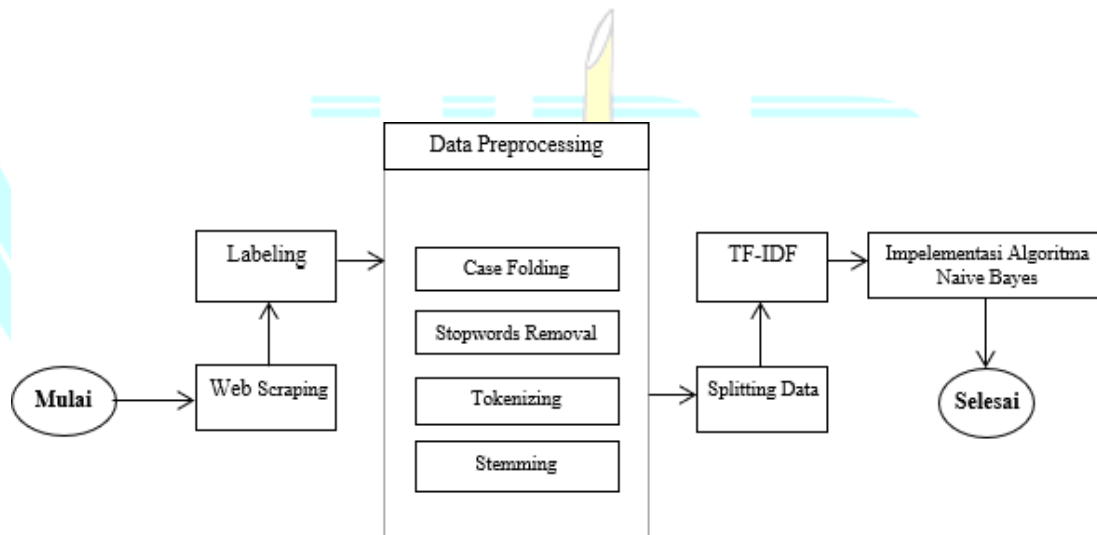
1. Menentukan sumber data, yaitu halaman aplikasi Jaket Boat pada Google Play Store.
2. Mengidentifikasi elemen HTML tempat ulasan ditampilkan.
3. Membuat skrip scraping untuk mengambil teks ulasan pengguna secara otomatis.
4. Menyimpan hasil ulasan ke dalam format CSV untuk memudahkan pemrosesan selanjutnya.

3.2.2 Alur Proses Penelitian

Secara umum, alur penelitian ini yang mencerminkan tahapan proses dari awal hingga akhir.

Tahapan tersebut meliputi :

1. Pengumpulan data ulasan dari Play Store.
2. Pelabelan data berdasarkan skor ulasan
3. Preprocessing teks untuk membersihkan dan menyiapkan data
4. Ekstraksi fitur menggunakan metode TF-IDF
5. Pembagian data menjadi data latih dan data uji
6. Penerapan algoritma Naïve Bayes
7. Evaluasi hasil klasifikasi menggunakan metrik performa.



Gambar 3. 2 Proses Penelitian

3.2.3 Preprocessing

Penelitian ini dilakukan melalui beberapa tahapan sebagai berikut:

1. Preprocessing Data

Preprocessing bertujuan untuk menghilangkan noise dan menyeragaman bentuk kata untuk mengurangi volume kosakata [Agustian, A., & Nurapriani, F. 2022]. Tahapan ini bertujuan untuk membersihkan dan

menyiapkan teks agar dapat diproses secara komputasional. Proses ini penting karena data mentah yang diperoleh dari scraping biasanya masih mengandung banyak noise, seperti tanda baca, angka, atau kata tidak bermakna. Dalam tahap ini terdapat beberapa langkah yang harus dilakukan seperti case folding, tokenizing, filtering, stemming

1. Pembersihan Data: Dari hasil scraping data, maka dilakukan tahapan cleaning data untuk menghapus beberapa karakter yang sering muncul Tokenisasi : memecah teks menjadi kata-kata individual [Irvandi, B. Irawan, and O. Nurdiawan].
2. Case Folding: Mengubah seluruh huruf menjadi huruf kecil untuk menyamakan representasi kata, misalnya “Bagus” dan “bagus” menjadi “bagus”.
3. Tokenisasi: Memecah teks ulasan menjadi kata-kata terpisah agar dapat diproses lebih lanjut.
4. Stopword Removal: Menghapus kata-kata umum (seperti "yang", "dan", "adalah") yang tidak memiliki makna penting dalam analisis sentimen.
5. Stemming: Mengubah kata menjadi bentuk dasar menggunakan algoritma sastrawi. Contohnya, “pembayaran”, “dibayar”, dan “bayar” akan dikembalikan ke bentuk dasar “bayar”.

2. Ekstraksi Fitur

Setelah teks dibersihkan, tahap selanjutnya adalah mengubah data teks menjadi bentuk numerik agar dapat digunakan dalam algoritma pembelajaran mesin. Dalam penelitian ini digunakan metode TF-IDF (Term Frequency – Inverse Document Frequency), yang mengukur seberapa penting sebuah kata terhadap suatu dokumen dalam kumpulan dokumen.

Rumus dasar:

1. **TF (Term Frequency)**: Mengukur seberapa sering kata muncul dalam dokumen.
2. **IDF (Inverse Document Frequency)**: Mengukur seberapa penting kata tersebut dalam seluruh korpus.

Kata-kata yang sering muncul dalam satu ulasan tetapi jarang muncul di seluruh dokumen akan memiliki nilai TF-IDF tinggi dan dianggap penting.

3. Penerapan Algoritma Naïve Bayes

Setelah data dalam bentuk vektor numerik, dilakukan klasifikasi menggunakan algoritma Naïve Bayes Multinomial. Algoritma ini dipilih karena kemampuannya yang baik dalam menangani data teks dan efisiensi dalam waktu komputasi.

Langkah-langkah:

- 1) Membagi data menjadi data latih dan data uji (80:20).
- 2) Melatih model Naïve Bayes menggunakan data latih yang telah dilabeli sentimen (positif/negatif).
- 3) Menggunakan model untuk mengklasifikasikan sentimen pada data uji.
- 4) Melakukan prediksi dan mencatat hasil klasifikasi.

3.3 Evaluasi dan Validasi

Untuk mengukur performa algoritma Naïve Bayes dalam klasifikasi sentimen, digunakan metrik evaluasi seperti:

1. Akurasi: Persentase prediksi yang benar terhadap total data uji.
2. Presisi: Ketepatan model dalam mengidentifikasi sentimen tertentu.
3. Recall: Kemampuan model dalam menangkap semua data yang sesuai dengan kategori tertentu.
4. F1-Score: Rata-rata harmonis antara presisi dan recall.