

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era digital yang terus mengalami percepatan kemajuan serta koneksi yang semakin luas, kebutuhan masyarakat terhadap layanan yang serba cepat, efisien, dan berbasis teknologi mengalami peningkatan yang sangat signifikan, terutama dalam sektor transportasi dan pariwisata. Traveloka, sebagai salah satu platform perjalanan daring terkemuka di kawasan Asia Tenggara, telah berhasil menjawab kebutuhan tersebut melalui penyediaan beragam layanan, mulai dari pemesanan tiket pesawat dan akomodasi hingga berbagai produk gaya hidup, yang semuanya dapat diakses dalam satu aplikasi terpadu. Dikutip dari (Databoks.katadata.co.id 2022, 16 Juni), Traveloka menempati posisi teratas sebagai situs perjalanan dengan trafik tertinggi di Indonesia, mencatatkan 7,2 juta kunjungan pada Maret 2022. Angka ini melonjak secara signifikan menjadi 22,3 juta kunjungan pada April 2024, menunjukkan adanya peningkatan eksponensial dalam jumlah pengguna dalam kurun waktu dua tahun terakhir. Tak hanya itu, aplikasi Traveloka juga telah diunduh lebih dari 50 juta kali di platform *Google Play Store*, mencerminkan penetrasi pasar yang luas serta tingginya kepercayaan konsumen terhadap mutu dan keandalan layanan yang ditawarkan.

Banyaknya pengguna yang memanfaatkan layanan ini turut menghasilkan volume ulasan dan umpan balik yang besar, yang mencakup apresiasi terhadap kemudahan layanan hingga kritik terhadap kendala teknis yang dialami. Ulasan-ulasan ini bukan hanya berperan sebagai indikator tidak langsung dari kepuasan pengguna, namun juga menjadi sumber informasi yang kaya untuk mengevaluasi kualitas layanan yang diberikan. Oleh karena itu, analisis terhadap sentimen dalam ulasan pengguna menjadi langkah penting untuk memahami secara lebih dalam persepsi masyarakat terhadap Traveloka. Hasil analisis ini dapat memberikan landasan yang kuat bagi perusahaan dalam merumuskan strategi untuk meningkatkan kualitas layanan dan kepuasan pelanggan secara berkesinambungan.

Seiring dengan meningkatnya jumlah ulasan yang masuk, dibutuhkan sistem berbasis otomatisasi yang mampu melakukan analisis sentimen dengan cepat, akurat, dan efisien. (Rivaldi dan Wismarini, 2024) menjelaskan bahwa analisis

sentimen digunakan untuk mengidentifikasi kecenderungan sikap dan ekspresi emosional yang disampaikan dalam bentuk teks. Pendekatan ini berperan dalam menggambarkan persepsi pengguna terhadap berbagai objek, seperti topik tertentu, produk, layanan, maupun suatu merek. Dalam penerapannya, salah satu metode klasifikasi yang banyak digunakan adalah *Naïve Bayes Classifier* (NBC). Metode ini memiliki keunggulan berupa proses penerapan yang sederhana serta kemampuan untuk tetap bekerja dengan baik pada data yang tidak lengkap dan atribut yang beragam. Selain itu, NBC memiliki ketahanan terhadap gangguan data (*noise*) dan mampu mempertahankan kinerja yang optimal ketika digunakan pada dataset berskala besar. Keunggulan tersebut menjadikan *Naïve Bayes Classifier* relevan dan efektif untuk berbagai permasalahan pengolahan teks, khususnya dalam analisis sentimen (Rivaldi & Wismarini, 2024).

Meski demikian, tantangan terbesar dalam klasifikasi berbasis teks adalah besarnya jumlah fitur yang dihasilkan dari proses ekstraksi, seperti tokenisasi dan pembobotan kata. Jumlah fitur yang berlebihan dapat meningkatkan kompleksitas model, menyebabkan *overfitting*, dan pada akhirnya mengurangi efektivitas serta efisiensi prediksi. Oleh karena itu, penerapan seleksi fitur menjadi langkah penting untuk memilih atribut yang benar-benar berpengaruh dalam proses klasifikasi. Salah satu metode seleksi fitur yang banyak digunakan dan terbukti memiliki kinerja baik adalah *Information Gain (IG)*. Hasil pemilihan fitur menunjukkan bahwa penggunaan sekitar 50% fitur dengan nilai *Information Gain* dan *Gain Ratio* tertinggi mampu menghasilkan tingkat akurasi yang lebih baik dibandingkan dengan metode penyaringan fitur lainnya (Rivaldi & Wismarini, 2024).

Penggunaan seleksi fitur *Information Gain (IG)* pada proses klasifikasi ulasan *Traveloka* dengan *Naïve Bayes Classifier (NBC)* didasarkan pada beberapa pertimbangan utama. Salah satunya, metode *Information Gain* mampu menekan dimensi data dengan menyaring fitur yang memiliki pengaruh kecil terhadap hasil klasifikasi, sehingga tingkat kompleksitas model dapat dikurangi sekaligus mempercepat proses komputasi. Hal ini sangat relevan mengingat ulasan di *Google Play Store* sering kali mengandung variasi bahasa yang luas, termasuk slang, emotikon, dan kalimat tidak terstruktur. Kedua, *Information Gain (IG)* meningkatkan akurasi klasifikasi dengan memprioritaskan fitur-fitur yang memiliki

korelasi kuat dengan label sentimen, seperti kata-kata yang secara eksplisit menunjukkan kepuasan atau keluhan pengguna. Selain itu, efektivitas pendekatan ini telah dibuktikan dalam berbagai penelitian terdahulu. Salah satunya ditunjukkan oleh studi yang dilakukan (Rita dan Pambudi, 2023), yang melaporkan bahwa penerapan seleksi fitur *Information Gain (IG)* pada klasifikasi ulasan aplikasi *Peduli Lindungi* mampu meningkatkan tingkat akurasi *Naïve Bayes Classifier (NBC)*.

Dengan menerapkan *Information Gain (IG)* sebagai langkah pra-pemrosesan sebelum menjalankan klasifikasi menggunakan algoritma *Naïve Bayes Classifier (NBC)*, diharapkan proses klasifikasi dapat berjalan lebih optimal. *Information Gain (IG)* berperan dalam menghapus fitur-fitur yang tidak relevan serta mengandung noise, yang dapat mengganggu proses pelatihan model dan menurunkan akurasi hasil prediksi. Pendekatan ini tidak hanya mampu meningkatkan ketepatan klasifikasi, tetapi juga mempercepat proses komputasi dan meningkatkan keterbacaan hasil model (Yuni & Santi, 2024). Penelitian ini bertujuan untuk mengevaluasi sejauh mana penerapan metode *Information Gain (IG)* dapat meningkatkan performa algoritma *Naïve Bayes Classifier (NBC)* dalam mengklasifikasikan sentimen ulasan pengguna terhadap layanan Traveloka. Harapannya, hasil dari penelitian ini dapat memberikan wawasan yang lebih komprehensif mengenai pentingnya seleksi fitur dalam pemodelan analisis sentimen, serta memberikan kontribusi dalam perancangan sistem klasifikasi yang lebih akurat, efisien, dan relevan dengan kebutuhan industri digital saat ini. Selain itu, temuan penelitian ini juga diharapkan dapat digunakan sebagai pijakan dalam pengambilan kebijakan strategis oleh pihak Traveloka, guna terus meningkatkan kualitas layanan dan kepuasan pengguna secara berkelanjutan di masa mendatang.

1.2 Rumusan Masalah

Rumusan masalah yang akan disusun dalam tugas akhir ini meliputi:

1. Bagaimana penerapan seleksi fitur *Information Gain (IG)* dapat mengidentifikasi fitur yang paling relevan dalam ulasan pengguna aplikasi Traveloka di *Google Play Store*?

2. Seberapa besar pengaruh seleksi fitur *Information Gain* (IG) terhadap akurasi, presisi, recall, dan *F1-Score* dalam klasifikasi sentimen ulasan Traveloka menggunakan *Naïve Bayes Classifier* (NBC)?
3. Seberapa besar perbandingan akurasi penerapan seleksi fitur *Information Gain* (IG) dengan tanpa menggunakan seleksi fitur?

1.3 Tujuan Masalah

Berdasarkan permasalahan yang telah diidentifikasi, tujuan dari penelitian ini adalah:

1. Mengevaluasi performa metode *Naïve Bayes Classifier* (NBC) setelah penerapan seleksi fitur *Information Gain* (IG) dengan tanpa menggunakan seleksi fitur.
2. Menganalisis pengaruh seleksi fitur *Information Gain* (IG) terhadap metrik evaluasi (akurasi, presisi, *recall*, dan *F1-Score*) dalam klasifikasi sentimen ulasan Traveloka menggunakan *Naïve Bayes Classifier* (NBC).
3. Mengetahui perbandingan akurasi setelah penerapan seleksi fitur *Information Gain* (IG) dengan tanpa menggunakan seleksi fitur.

1.4 Manfaat Penelitian

Adapun manfaat dari tercapainya penelitian ini yaitu:

1. Bagi peneliti, hasil ini berguna sebagai materi untuk menyusun tugas akhir dan untuk mengembangkan pengetahuan yang telah diperoleh selama studi di perguruan tinggi.
2. Memberikan kontribusi pengetahuan di bidang akademik sehingga dapat menambah pemahaman mengenai implementasi analisis sentimen menggunakan metode *Naïve Bayes Classifier* (NBC) dengan seleksi fitur *Information Gain* (IG).

1.5 Batasan Masalah

Adapun beberapa batasan masalah yang diangkat dalam melakukan penelitian ini, antara lain:

1. Data yang digunakan terbatas pada ulasan pengguna Traveloka yang diperoleh dari *Google Play store* berjumlah 500 data. Dengan data latih sebanyak 400 data dan data uji sebanyak 100 data.

2. Metode yang digunakan adalah *Naïve Bayes Classifier* (NBC) tanpa perbandingan dengan metode lain dan menggunakan fitur seleksi *Information Gain* (IG).
3. Dataset yang diambil adalah kumpulan ulasan yang berkaitan dengan topik yang sedang terjadi.
4. Klasifikasi sentimen pengguna yang dilakukan berdasarkan sentiment positif, netral, negatif.
5. Dataset yang diambil tidak melibatkan gambar, video, atau tautan.
6. Data ulasan yang akan di analisis menggunakan bahasa Indonesia dan orang Indonesia .
7. Total data ulasan yang digunakan dalam penelitian diperoleh mulai tanggal 29 mei 2025 sampai 17 juni 2025.
8. Perangkat lunak yang digunakan sebagai media analisis adalah *Google Colab* dan bahasa pemrograman *python*.

