

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Dalam dua dekade terakhir, percepatan arus informasi dan perkembangan platform digital telah memperluas ruang partisipasi publik sehingga opini masyarakat terhadap industri hiburan digital dapat terbentuk dan tersebar secara lebih cepat (Prasetyo, 2023). Di antara berbagai bentuk hiburan digital, industri *game* menonjol karena pertumbuhannya yang sangat cepat. *Game* berbasis perangkat *mobile* telah mencatat lonjakan popularitas yang luar biasa, menarik perhatian jutaan pengguna (Andriansyah dkk., 2023). Salah satu *game* lokal yang berhasil menarik perhatian masyarakat Indonesia adalah Bus Simulator Indonesia (BUSSID).

Bus Simulator Indonesia (BUSSID), yang dikembangkan oleh Maleo, merupakan salah satu contoh keberhasilan *game* lokal yang mampu bersaing di pasar *mobile game*. *Game* ini menyajikan pengalaman simulasi mengemudi bus dengan nuansa khas Indonesia, seperti desain kendaraan yang dapat dikustomisasi, peta jalanan lokal, serta berbagai elemen budaya populer seperti klakson “*Om Telolet Om*”. Sejak dirilis, BUSSID telah meraih lebih dari seratus juta unduhan di Google Play Store dan mendapat jutaan ulasan dari pengguna (Prastya, 2024). Banyaknya ulasan ini menjadi indikator penting akan tingginya interaksi dan perhatian dari para pemain.

Ulasan pengguna merupakan salah satu bentuk umpan balik langsung yang mencerminkan kepuasan, kritik, maupun harapan pemain terhadap sebuah aplikasi, termasuk *game*. Dalam konteks *game* Bus Simulator Indonesia, ulasan yang diberikan pengguna mencakup berbagai aspek, seperti kualitas grafis, fitur *gameplay*, kestabilan aplikasi, hingga pengalaman bermain secara keseluruhan. Meskipun demikian, banyaknya jumlah ulasan yang bersifat teks bebas menjadikannya sulit untuk dianalisis secara manual. Selain itu, ragam ekspresi dalam bahasa, gaya penulisan, serta nada emosional yang berbeda-beda membuat proses interpretasi menjadi semakin kompleks (Mursidah dkk., 2025).

Meskipun ulasan pengguna dapat menjadi sumber informasi yang sangat berharga, jumlahnya yang sangat banyak dan bentuknya yang tidak terstruktur menjadikan proses analisis secara manual menjadi tidak efisien. Sebagian besar ulasan ditulis dalam bentuk teks bebas dengan gaya penulisan yang sangat beragam, mulai dari penggunaan bahasa formal hingga bahasa sehari-hari yang disertai dengan singkatan, emotikon, dan ekspresi subjektif. Selain itu, nada emosional yang disampaikan pengguna dalam ulasan pun bervariasi, mulai dari keluhan, pujian, hingga saran, yang tidak selalu mudah dikenali secara langsung. Kompleksitas ini menyebabkan interpretasi terhadap sentimen dalam ulasan menjadi tantangan tersendiri (Indrayanto dkk., 2023). Oleh karena itu, diperlukan suatu pendekatan berbasis teknologi untuk metode klasifikasi teks, seperti algoritma *Naïve Bayes*, sebagai alat bantu dalam pengelompokan ulasan berdasarkan kecenderungan sentimen secara lebih sistematis dan akurat.

*Naïve bayes* merupakan sebuah teknik klasifikasi yang berbasis pada prinsip Teorema Bayes. Algoritma ini bekerja dengan menghitung probabilitas masing-masing kelas terhadap fitur-fitur dalam data. Algoritma ini disebut dengan kata "Naïve" karena metode ini mengasumsikan bahwa setiap fitur bersifat independen satu sama lain, meskipun dalam kenyataannya, fitur-fitur tersebut kadang kala saling berkaitan. *Naïve Bayes* terbukti efektif dan efisien dalam berbagai tugas klasifikasi teks, termasuk dalam analisis ulasan pengguna. Keunggulan utama dari algoritma ini adalah kecepatan dalam proses pelatihan data, kemudahan implementasi, serta performa yang cukup baik bahkan pada *dataset* yang besar dan tidak terstruktur seperti ulasan aplikasi di platform digital. Meskipun *Naïve Bayes* cepat dan mudah dalam klasifikasi, hasilnya sangat dipengaruhi oleh bagaimana data teks direpresentasikan (Azzami dkk., 2025). Salah satu metode yang efektif untuk merepresentasikan data teks adalah *Term Frequency-Inverse Document Frequency* (TF-IDF), karena mampu menilai pentingnya kata dalam konteks dokumen dan keseluruhan korpus.

*Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan teknik yang digunakan untuk merepresentasikan teks dalam bentuk numerik dengan memberikan bobot pada setiap kata. Bobot ini dihitung berdasarkan dua komponen

utama, yaitu frekuensi kemunculan kata dalam sebuah dokumen (*term frequency*) dan seberapa jarang kata tersebut muncul di seluruh kumpulan dokumen (*inverse document frequency*). Dengan pendekatan ini, TF-IDF membantu menekankan kata-kata yang lebih penting dan relevan dalam sebuah teks dan memudahkan proses analisis selanjutnya. Oleh karena itu, TF-IDF dan *Naïve Bayes* dipilih sebagai pendekatan utama dalam penelitian ini untuk mengklasifikasikan sentimen pengguna terhadap *game* Bus Simulator Indonesia.

Beberapa penelitian telah membuktikan bahwa kontribusi signifikan dari metode *Naïve Bayes* dalam analisis sentimen. Susanto dkk., (2021) menunjukkan bahwa *Naïve Bayes* efektif dalam menganalisis sentimen data X (Twitter) terkait pemilihan kepala daerah yang merupakan konteks politik. Maharani dan Triayudi (2022) menemukan bahwa *Naïve Bayes* lebih unggul dibandingkan *K-Nearest Neighbor* (K-NN) dalam mengklasifikasikan sentimen pengguna terhadap layanan pembayaran digital seperti GOPAY, DANA, dan ShopeePay. Sementara itu, Ramadhan dkk. (2022) membuktikan bahwa *Naïve Bayes* memiliki performa tinggi dalam klasifikasi sentimen ulasan *e-commerce* dengan hasil akurasi dan ukuran performa model klasifikasi yang sangat baik.

## 1.2. Rumusan Masalah

Berdasarkan paparan latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana karakteristik ulasan pengguna di *Google Play Store* terhadap *game* Bus Simulator Indonesia ditinjau dari tiga jenis sentimen utama yaitu positif, negatif, dan netral.
2. Bagaimana algoritma TF-IDF dan *Naïve Bayes* dapat digunakan untuk mengklasifikasi dan menganalisis sentimen ulasan pengguna terhadap *game* Bus Simulator Indonesia, ditinjau dari segi akurasi (*accuracy*), presisi (*precision*), *recall*, dan *F1-score*.

### 1.3. Batasan Penelitian

Untuk memastikan penelitian ini mencapai target dan tujuan yang diharapkan, tantangan yang dihadapi saat ini terbatas pada beberapa aspek berikut.

1. Data yang digunakan terbatas pada ulasan pengguna *game* Bus Simulator Indonesia (BUSSID) dari platform *Google Play Store* antara tanggal 31 Maret 2025 hingga 14 April 2025.
2. Ulasan sentimen yang dianalisis dalam penelitian ini terbatas pada ulasan yang berbahasa Indonesia.
3. Pemrosesan data sentimen dilakukan menggunakan bahasa pemrograman *Python* dengan memanfaatkan platform *Google Colab*.
4. Evaluasi dan validasi model dilakukan menggunakan *confusion matrix* untuk mengukur kinerja klasifikasi.
5. Analisis sentimen dalam penelitian ini hanya diklasifikasikan ke dalam tiga kategori, yaitu positif, negatif, atau netral.
6. Penentuan kategori dilakukan dengan menggunakan *indonesia sentiment lexicon*.
7. Data ulasan yang digunakan dalam penelitian ini berjumlah 2.000.

### 1.4. Tujuan Penelitian

Tujuan yang hendak dicapai dalam penelitian ini adalah:

1. untuk mengidentifikasi sentimen pengguna terhadap aplikasi *game* Bus Simulator Indonesia berdasarkan ulasan yang diperoleh dari *Google Play Store*.
2. untuk mengetahui kehandalan metode TF-IDF dan *Naïve Bayes* dalam mengklasifikasikan sentimen pengguna ke dalam kategori positif, negatif, dan netral berdasarkan nilai akurasi, presisi, dan *recall*.

### 1.5. Manfaat Penelitian

Dengan penelitian ini, manfaat yang dapat diperoleh adalah sebagai berikut.

1. Penelitian ini diharapkan memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya dalam bidang *text mining* dan *machine learning*. Dengan menerapkan algoritma *Naïve Bayes* untuk analisis sentimen.
2. Penelitian ini dapat menjadi referensi dan acuan bagi penelitian-penelitian selanjutnya yang berfokus pada analisis ulasan pengguna atau data teks lainnya dalam konteks aplikasi atau *game*.

