

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah memberikan dampak signifikan terhadap kehidupan masyarakat modern, termasuk dalam bentuk kemunculan berbagai platform media sosial yang menjadi ruang diskusi dan interaksi publik. Platform X sebagai salah satu media sosial yang populer di Indonesia digunakan oleh jutaan pengguna untuk berbagi opini, pengalaman, serta memberikan ulasan terhadap berbagai topik, termasuk game. Salah satu game yang mendapatkan perhatian besar dari kalangan remaja hingga dewasa adalah Roblox. Game ini memiliki basis komunitas aktif yang sering memberikan respons positif maupun negatif di media sosial, sehingga memunculkan fenomena menarik untuk dianalisis dari sudut pandang opini publik.

Analisis sentimen merupakan bagian dari Natural Language Processing (NLP), yaitu cabang ilmu yang mempelajari bagaimana komputer memahami dan memproses bahasa manusia. NLP memungkinkan mesin untuk mengubah teks yang tidak terstruktur menjadi informasi yang dapat dianalisis secara otomatis. Dalam konteks analisis sentimen, proses ini dilakukan melalui tahapan seperti tokenisasi, stemming, pembobotan kata, hingga klasifikasi untuk menentukan apakah suatu opini bersifat positif, negatif, atau netral. Dengan kemampuan ini, sistem dapat menangkap ekspresi emosi dan persepsi pengguna secara lebih efisien dibandingkan metode manual. Oleh karena itu, analisis sentimen menjadi alat yang penting untuk mempelajari perilaku pengguna, terutama pada platform media sosial yang menghasilkan data dalam jumlah besar dan bersifat real-time.

Perkembangan penelitian di bidang analisis sentimen menunjukkan bahwa metode berbasis pembelajaran mesin, khususnya algoritma Naïve Bayes, mampu mengolah opini pengguna dalam jumlah besar secara efisien dan cukup akurat. Pada konteks ulasan aplikasi e-commerce seperti Shopee dan Lazada, proses analisis sentimen dilakukan melalui tahapan pra-pemrosesan teks (cleaning, case

folding, tokenizing, stopword removal, normalisasi, dan stemming), kemudian dilanjutkan dengan pembobotan kata menggunakan Term Frequency-Inverse Document Frequency (TF-IDF), sebelum akhirnya diklasifikasikan menggunakan algoritma Naive Bayes dan K-Nearest Neighbor [1]. Hasil penelitian tersebut menunjukkan bahwa Naive Bayes menghasilkan akurasi hingga sekitar 82–84% dan terbukti lebih stabil dibanding KNN, sehingga direkomendasikan sebagai algoritma yang tepat untuk klasifikasi sentimen pada data ulasan pengguna berskala besar. Pendekatan serupa dapat diadaptasi untuk menganalisis opini pengguna terhadap game Roblox di Platform X, dengan penyesuaian pada karakteristik bahasa dan konteks diskusi di media sosial.

Efektivitas Naive Bayes dalam klasifikasi teks juga diperkuat oleh penelitian lain pada domain hiburan dan aplikasi digital. Studi terhadap 50.000 ulasan film pada dataset IMDB melaporkan bahwa Naive Bayes mampu mencapai akurasi 86,25% dengan nilai precision, recall, dan F1-score masing-masing 86%, yang menunjukkan keseimbangan kinerja dalam membedakan sentimen positif dan negatif pada ulasan yang beragam [2]. Di sisi lain, kombinasi Multinomial Naive Bayes dengan pendekatan Design Thinking pada analisis sentimen pengguna terhadap antarmuka e-learning X terbukti efektif meningkatkan sentimen positif dan kepuasan pengguna setelah perbaikan antarmuka dilakukan berdasarkan hasil klasifikasi sentimen tersebut[3]. Temuan-temuan ini mengindikasikan bahwa penerapan Multinomial Naive Bayes pada penelitian analisis sentimen opini pengguna Platform X terhadap game Roblox berpotensi tidak hanya menghasilkan pemetaan sentimen yang andal, tetapi juga menyediakan dasar yang kuat bagi pengembang maupun pemangku kepentingan untuk memahami persepsi pengguna dan merumuskan strategi peningkatan pengalaman bermain maupun interaksi komunitas secara lebih terarah.

1.2 Rumusan Masalah

1. Apa kecenderungan sentimen pengguna Platform X terhadap Game Roblox di Indonesia ?
2. Metode Naive Bayes(Multinomial Naive Bayes, Bernoulli Naive Bayes,

dan Gaussian Naive Bayes) manakah yang menunjukkan performa terbaik berdasarkan akurasi, presisi, recall, F1-score, dan ROC?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan tidak menyimpang dari tujuan yang telah ditetapkan, maka diperlukan batasan masalah dalam pelaksanaannya. Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini berasal dari hasil scraping komentar pengguna pada Platform X (Twitter) menggunakan bantuan Grox AI. Pengambilan data dibatasi hanya pada komentar yang berkaitan dengan Game Roblox di Indonesia, sehingga tidak mencakup platform media sosial lain seperti Instagram, Facebook, atau Reddit. Data hanya menggunakan kata kunci "roblox indonesia", sehingga komentar yang tidak secara eksplisit mengandung atau berkaitan dengan kata kunci tersebut tidak dimasukkan ke dalam dataset penelitian, meskipun masih memiliki relevansi secara tidak langsung.
2. Penelitian ini hanya membagi sentimen pengguna ke dalam dua kelas, yaitu negatif, dan positif. Kategori sentimen lain seperti sangat positif atau sangat negatif tidak digunakan untuk menjaga konsistensi dan kesederhanaan proses klasifikasi.
3. Transformasi fitur teks dalam penelitian ini dibatasi pada pendekatan utama, yaitu TF-IDF. Metode representasi teks lain seperti Bag of Words murni, FastText, atau Transformer-based embedding tidak digunakan dalam penelitian ini.
4. Evaluasi performa model dalam penelitian ini dibatasi pada metrik akurasi, presisi, recall, dan F1-score. Metrik lain seperti confusion cost matrix tidak digunakan sebagai indikator utama penilaian.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis performa seluruh turunan algoritma Naive Bayes, yaitu Multinomial Naive Bayes, Bernoulli Naive Bayes, dan Gaussian Naive Bayes, dalam melakukan klasifikasi sentimen terhadap komentar pengguna Platform X.
2. Menentukan turunan algoritma Naive Bayes dengan pendekatan ekstraksi fitur terbaik berdasarkan hasil evaluasi performa sebagai model paling optimal dalam klasifikasi sentimen pengguna Platform X terhadap Game Roblox di Indonesia.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis, sebagai berikut:

1. Hasil penelitian ini dapat menjadi referensi akademik mengenai pengaruh penggunaan metode ekstraksi fitur TF-IDF terhadap performa algoritma Naive Bayes dalam analisis sentimen media sosial.
2. Penelitian ini diharapkan mampu memberikan gambaran mengenai karakteristik data sentimen pengguna Platform X di Indonesia, khususnya pada topik Game Roblox, sehingga dapat dimanfaatkan dalam penelitian sejenis di masa mendatang.

1.6 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun untuk memberikan gambaran secara umum mengenai isi dari setiap bab yang terdapat dalam laporan penelitian. Adapun sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN, berisi latar belakang penelitian yang menjelaskan alasan dan urgensi dilakukannya analisis sentimen terhadap komentar pengguna

Platform X terkait game Roblox di Indonesia. Selain itu, bab ini memuat rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan yang memberikan gambaran umum mengenai struktur penyusunan skripsi.

BAB II TINJAUAN PUSTAKA, membahas landasan teori yang relevan dengan penelitian, meliputi konsep analisis sentimen, text mining, pra-pemrosesan teks, metode TF-IDF, algoritma Naive Bayes beserta turunannya, serta metrik evaluasi kinerja model klasifikasi. Selain itu, bab ini juga memuat tinjauan penelitian terdahulu yang berkaitan sebagai dasar pembandingan dan penguat posisi penelitian yang dilakukan.

BAB III METODE PENELITIAN, menjelaskan metode dan tahapan penelitian yang digunakan secara sistematis, mulai dari pengumpulan data komentar pengguna Platform X melalui proses scraping, pelabelan data sentimen, pra-pemrosesan teks, pembagian data menjadi data latih dan data uji, transformasi data menggunakan metode TF-IDF, seleksi fitur, penerapan algoritma Naive Bayes, hingga proses evaluasi dan penentuan model terbaik.

BAB IV HASIL DAN PEMBAHASAN, menyajikan hasil penelitian yang diperoleh dari setiap tahapan metode yang telah diterapkan. Pembahasan meliputi hasil pengumpulan data, karakteristik dataset, hasil pra-pemrosesan teks, distribusi sentimen, hasil transformasi fitur TF-IDF, kinerja masing-masing model Naive Bayes berdasarkan metrik akurasi, presisi, recall, dan F1-score, serta analisis perbandingan performa model.

BAB V PENUTUP, berisi kesimpulan yang diperoleh berdasarkan hasil dan pembahasan penelitian, serta saran yang dapat dijadikan acuan untuk pengembangan dan penelitian selanjutnya terkait analisis sentimen atau penerapan metode klasifikasi teks yang lebih lanjut.