

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi membawa perubahan pada dunia pendidikan melalui pembelajaran digital yang lebih fleksibel. Salah satu inovasinya adalah Ruangguru yang didirikan tahun 2014 oleh Belva Devara dan Iman Usman untuk menyediakan layanan pembelajaran online dan corporate e-learning (Aprianti et al., 2025). Perkembangan teknologi mendorong perubahan dalam dunia pendidikan melalui hadirnya platform pembelajaran digital yang fleksibel. Salah satu contohnya adalah Ruangguru yang didirikan pada tahun 2014 oleh Belva Devara dan Iman Usman. Platform ini menyediakan layanan pembelajaran online dan corporate e-learning, serta memiliki 38 juta pengguna dengan sekitar 200 juta akses per bulan (Muhammad Nafis et al., 2025).

Tingkat kepuasan pengguna terhadap Ruangguru di Google Play Store dapat menjadi bahan evaluasi bagi perusahaan dalam menentukan tujuan dan memantau kinerja layanan yang diberikan (Kamil et al., 2025). Pengguna biasanya menyampaikan pendapat melalui kolom komentar di Google Play Store untuk memberikan kritik maupun saran terhadap suatu aplikasi (Kurniawan et al., 2025). Seperti aplikasi teknologi lainnya, kualitas layanan Ruangguru dipengaruhi oleh umpan balik dan kepuasan pengguna. Komentar serta ulasan di Google Play Store dapat digunakan untuk menganalisis persepsi dan sentimen pengguna terhadap layanan tersebut (Sofia & Sekarpuji, 2025). Informasi tersebut memberikan gambaran mengenai pengalaman pengguna dalam menggunakan aplikasi. Hasil analisis komentar juga dapat dimanfaatkan sebagai dasar evaluasi dan perbaikan layanan secara berkelanjutan.

Dalam penelitian ini, analisis sentimen dilakukan melalui teknik crawling dan *scraping* untuk mengumpulkan data ulasan pengguna di Google Play Store. Penelitian bertujuan menganalisis sentimen positif, negatif, dan netral pada ulasan pengguna Ruangguru menggunakan algoritma Naïve Bayes. Hasil analisis diharapkan dapat menjadi dasar pengembangan dan peningkatan layanan di masa mendatang (Damayanti et al., 2025). Naïve Bayes merupakan algoritma yang sering

digunakan dalam analisis sentimen dan didasarkan pada Teorema Bayes. Algoritma ini dipilih karena sederhana, efisien untuk dataset besar, serta memiliki kinerja yang baik dalam klasifikasi teks (Ramadhan As'ari et al., 2025).

Untuk meningkatkan kualitas representasi data sebelum proses klasifikasi dilakukan, data teks diolah menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), yaitu teknik statistik untuk menghitung frekuensi sekaligus bobot kepentingan suatu kata dalam dokumen. Metode ini dikenal sederhana, efisien, dan akurat, serta semakin sering sebuah kata muncul, semakin besar tingkat kepentingannya dalam teks (Nurzaman et al., 2024). Apabila suatu kata sering muncul pada satu dokumen tetapi jarang ditemukan pada dokumen lain, maka kata tersebut dianggap memiliki tingkat kepentingan yang tinggi pada dokumen tersebut (Aziz et al., 2024)

Karena data berbentuk teks menghasilkan banyak fitur, diperlukan reduksi dimensi fitur untuk menyaring atribut yang paling penting. Pendekatan *principal component analysis* (PCA) digunakan untuk mereduksi dimensi data agar lebih sederhana namun tetap mempertahankan informasi penting (Lestari et al., 2025). Reduksi dimensi pada data dapat dilakukan dengan mengidentifikasi arah utama penyebaran data yang disebut komponen utama (*principal component*). Metode PCA menganggap bahwa arah dengan varians terbesar merupakan informasi yang paling penting dalam data. Besarnya varians yang mampu dipertahankan oleh setiap komponen utama diukur menggunakan nilai eigen (Susanti et al., 2023).

Dalam penelitian ini, PCA diterapkan setelah pembobotan fitur menggunakan TF-IDF untuk mengatasi tingginya dimensi data teks yang dapat menghambat kinerja model. Temuan sebelumnya menunjukkan bahwa reduksi dimensi dengan PCA mampu mempertahankan informasi penting dari fitur TF-IDF sekaligus meningkatkan efisiensi proses klasifikasi berbasis ANN (Afrad, 2024). Selain itu, penelitian lain yang menerapkan PCA pada matriks TF-IDF untuk teks pendek dengan mereduksi ribuan fitur menjadi sekitar 500 komponen juga membuktikan bahwa penyederhanaan tersebut berhasil menurunkan kompleksitas model tanpa menurunkan kualitas prediksi (Khan et al., 2021). Temuan-temuan ini menegaskan pentingnya penggunaan PCA dalam penelitian ini sebagai langkah strategis untuk menghasilkan model yang lebih ringkas, cepat, dan tetap akurat.

Penelitian dengan Judul “Analisis Sentimen Penggunaan Aplikasi KineMaster Menggunakan Metode Naïve Bayes”. Dari total 1.200 ulasan aplikasi Kinemaster, setelah melalui tahap *preprocessing* diperoleh 1.031 data bersih. Dataset tersebut kemudian dibagi menggunakan skema 80:20, di mana 80% digunakan sebagai data latih dan 20% sebagai data uji. Hasil pengujian model menunjukkan tingkat *accuracy* sebesar 85%. Pada tahap evaluasi, diperoleh nilai *precision* sebesar 82%, *recall* 74%, *F1-score* 78%, serta *support* untuk kategori positif mencapai 88%, lebih tinggi dibandingkan kategori negatif yang hanya sebesar 78%. Walaupun performa implementasi sudah cukup baik, masih terdapat ruang untuk peningkatan, baik dari sisi *preprocessing* data maupun pemilihan algoritma yang lebih sesuai dengan karakteristik dataset (Kevin et al., 2024).

Penelitian “*The Impact of Principal Component Analysis Dimensionality Reduction on Sentimen Classification Performance Using Support Vector Machine*” berdasarkan penelitian tersebut, penerapan PCA terbukti mampu meningkatkan performa klasifikasi sentimen menggunakan algoritma SVM. Setelah penerapan PCA, nilai *accuracy* meningkat dari 72,65% menjadi 73,20%, *precision* dari 71,58% menjadi 72,24%, *recall* dari 71,77% menjadi 72,66%, dan *F1-score* dari 71,56% menjadi 72,32%. Hasil ini menunjukkan bahwa PCA efektif dalam menyederhanakan data teks berdimensi tinggi, mengurangi kompleksitas fitur, serta meningkatkan efisiensi dan kinerja model klasifikasi sentimen.

Berdasarkan penelitian analisis sentimen sebelumnya terhadap aplikasi KineMaster menggunakan metode Naïve Bayes, diperoleh hasil klasifikasi yang cukup baik dengan *accuracy* sebesar 85%. Namun, penelitian tersebut masih memiliki keterbatasan karena belum menerapkan teknik reduksi dimensi untuk mengatasi tingginya kompleksitas fitur pada data teks, serta belum menganalisis pengaruh kombinasi metode pembobotan fitur dan reduksi dimensi terhadap performa klasifikasi. Di sisi lain, penelitian mengenai penerapan Principal Component Analysis (PCA) pada klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM) menunjukkan bahwa PCA mampu meningkatkan performa model dengan menyederhanakan data berdimensi tinggi. Akan tetapi, penelitian tersebut hanya berfokus pada algoritma SVM sehingga efektivitas PCA pada algoritma Gaussian Naïve Bayes belum diketahui secara jelas.

Berdasarkan permasalahan tersebut, penelitian ini memiliki perbedaan dengan penelitian terdahulu karena mengombinasikan metode TF-IDF, PCA, dan Gaussian Naïve Bayes dalam analisis sentimen ulasan aplikasi Ruangguru. Selain itu, penelitian ini tidak hanya melakukan klasifikasi sentimen, tetapi juga menganalisis pengaruh reduksi dimensi PCA terhadap performa Gaussian Naïve Bayes pada data ulasan berdimensi tinggi. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi berupa penerapan kombinasi metode yang belum banyak digunakan pada analisis sentimen aplikasi pembelajaran serta menjadi pembanding efektivitas PCA pada algoritma Gaussian Naïve Bayes.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Permasalahan utama dalam penelitian ini berfokus pada pengolahan ulasan pengguna aplikasi Ruangguru di *Google Play Store* yang berjumlah besar serta memiliki variasi penggunaan bahasa. Tantangan yang dihadapi adalah mengklasifikasikan ulasan tersebut ke dalam tiga kategori sentimen, yakni positif, negatif, dan netral. Metode TF-IDF sering kali menghasilkan jumlah fitur yang sangat banyak, sehingga diperlukan penerapan reduksi dimensi menggunakan PCA agar proses klasifikasi dengan algoritma *Naïve Bayes* dapat berjalan lebih efisien dan menghasilkan *accuracy* yang lebih baik.

1.2.2 Metode Usulan

Dalam penelitian ini diusulkan metode *Naïve Bayes* sebagai algoritma klasifikasi data, metode TF-IDF sebagai teknik pembobotan kata, serta metode PCA untuk reduksi dimensi fitur. Pendekatan ini digunakan untuk mengolah ulasan pengguna aplikasi Ruangguru dari *Google Play Store* agar dapat diklasifikasikan ke dalam tiga sentimen, yaitu positif, negatif, dan netral dengan hasil akhir dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

1.3 Pertanyaan Penelitian

Berdasarkan permasalahan tersebut terdapat pertanyaan yaitu:

1. Bagaimana hasil penerapan analisis sentimen terhadap ulasan pengguna aplikasi Ruangguru di *Google Play Store* menggunakan algoritma *Naïve Bayes*?
2. Apakah penerapan metode reduksi dimensi menggunakan PCA (*Principal Component Analysis*) dapat meningkatkan *accuracy* model *Naïve Bayes*?

dibandingkan dengan model tanpa PCA dalam analisis sentimen terhadap ulasan pengguna aplikasi Ruangguru?

3. Bagaimana kinerja model *Naïve Bayes* berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Dalam penelitian ini terdapat tujuan yang ingin dicapai yaitu:

1. Mengetahui hasil analisis sentimen terhadap ulasan pengguna aplikasi Ruangguru menggunakan algoritma *Naïve Bayes*.
2. Menganalisis perbandingan kinerja model *Naïve Bayes* sebelum dan sesudah penerapan PCA (Principal Component Analysis).
3. Mengevaluasi kinerja model klasifikasi sentimen menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

1.4.2 Manfaat

Penelitian ini memiliki manfaat yaitu salah satunya memperkaya referensi dalam kajian analisis sentimen melalui penerapan TF-IDF, PCA, dan *Naïve Bayes*. Dari sisi praktis, hasil penelitian memberikan gambaran mengenai persepsi pengguna terhadap aplikasi Ruangguru sehingga dapat dijadikan masukan bagi pengembang dalam meningkatkan kualitas layanan.

1.5 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data yang digunakan hanya berasal dari ulasan pengguna aplikasi Ruangguru di *Google Play Store*.
2. Analisis sentimen difokuskan pada 3 kelas yaitu positif, negatif, dan netral
3. Metode yang digunakan hanya meliputi TF-IDF untuk ekstraksi fitur, PCA untuk reduksi dimensi, dan *Naïve Bayes* sebagai algoritma klasifikasi.
4. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.
5. Data yang digunakan dalam penelitian ini berasal dari 5.000 ulasan aplikasi Ruangguru di *Google Play Store*. Setelah dilakukan filtering spam, penghapusan data duplikasi, preprocessing, dan penghapusan data kosong, diperoleh 4.958 data akhir yang digunakan dalam proses penelitian.

1.6 Sistematika Penulisan

Dalam penelitian skripsi penulisan disusun secara sistematis dan dibagi menjadi beberapa bab, yaitu sebagai berikut :

BAB I. PENDAHULUAN

Bab ini berisi uraian awal yang menjadi dasar dari penelitian, meliputi latar belakang yang menjelaskan alasan dan urgensi topik yang diangkat, rumusan masalah yang merangkum inti persoalan yang akan dikaji, serta batasan masalah yang berfungsi untuk memperjelas ruang lingkup penelitian. Selain itu, bab ini juga mencantumkan tujuan penelitian yang ingin dicapai, baik secara umum maupun khusus, serta manfaat penelitian yang dapat memberikan kontribusi secara teoritis maupun praktis. Sebagai penutup, disajikan sistematika penulisan yang menggambarkan susunan bab demi bab, sehingga pembaca dapat memahami alur penelitian secara menyeluruh.

BAB II. TINJAUAN PUSTAKA

Bab ini memaparkan teori-teori pendukung penelitian, seperti konsep dasar analisis sentimen, metode *Naïve Bayes*, TF-IDF, PCA, serta penelitian-penelitian terdahulu yang relevan.

BAB III. METODE PENELITIAN

Bab ini menyajikan rancangan penelitian yang menjadi kerangka dasar dalam pelaksanaan analisis sentimen. Di dalamnya dijelaskan mengenai sumber data yang digunakan beserta alasan pemilihannya, serta teknik pengumpulan data yang diterapkan untuk memperoleh data ulasan secara sistematis. Selanjutnya, dipaparkan tahapan *preprocessing* yang mencakup berbagai langkah pengolahan teks agar data lebih bersih dan siap dianalisis. Bagian ini juga menguraikan metode analisis yang digunakan, meliputi TF-IDF sebagai teknik pembobotan kata dan ekstraksi fitur, PCA sebagai metode reduksi dimensi, serta algoritma *Naïve Bayes* untuk proses klasifikasi sentimen. Selain itu, data akan diproses dengan melakukan beberapa tahapan, diagram sistem, dan skenario pengujian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian serta pembahasan dari setiap tahapan yang telah dilakukan dalam analisis sentimen terhadap ulasan pengguna aplikasi Ruang-

guru. Proses penelitian diawali dengan pengolahan data melalui tahap *preprocessing* untuk menghasilkan data yang lebih terstruktur. Selanjutnya dilakukan pelabelan sentimen untuk mengklasifikasikan ulasan ke dalam kategori positif, negatif, dan netral. Data yang telah diproses kemudian diubah ke dalam bentuk numerik menggunakan metode TF-IDF dan direduksi dimensinya menggunakan *principal component analysis* (PCA). Untuk mengatasi ketidakseimbangan data, digunakan metode SMOTE, kemudian dilakukan proses klasifikasi menggunakan algoritma Naïve Bayes. Pengujian model dilakukan menggunakan metode 10-fold cross validation, sehingga diperoleh hasil evaluasi berupa nilai *accuracy*, *precision*, *recall*, dan *F1-score*. yang digunakan untuk menganalisis kinerja model secara keseluruhan.

BAB V PENUTUP

Bab ini merupakan bagian penutup dari penelitian yang berisi rangkuman hasil penelitian serta saran untuk pengembangan penelitian selanjutnya. Pada bagian ini dijelaskan mengenai kesimpulan yang diperoleh berdasarkan hasil analisis dan pengujian model yang telah dilakukan. Selain itu, disampaikan pula saran yang dapat digunakan sebagai bahan pertimbangan untuk penelitian selanjutnya agar dapat meningkatkan kualitas dan kinerja model analisis sentimen.