

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital yang pesat telah mendorong munculnya berbagai platform pembelajaran daring (*e-learning*), salah satunya adalah aplikasi Ruangguru. Aplikasi ini menyediakan layanan pendidikan berbasis teknologi yang memungkinkan siswa untuk belajar secara mandiri melalui video interaktif, latihan soal, dan fitur bimbingan belajar daring. Dengan meningkatnya jumlah pengguna dari tahun ke tahun, penting untuk memahami persepsi pengguna terhadap kualitas layanan yang diberikan Ruangguru melalui ulasan yang mereka tulis di berbagai platform digital.

Relevansi analisis sentimen dalam konteks aplikasi pendidikan juga telah dibahas oleh Pangaribuan dan Subakti [1], yang meneliti sistem informasi akademik berbasis web sebagai sarana pembelajaran digital. Ulasan pengguna yang tersedia di berbagai platform seperti Google Play Store, Twitter (X), Instagram, TikTok, dan YouTube mengandung informasi berharga terkait pengalaman serta tingkat kepuasan pengguna terhadap layanan yang digunakan. Namun, jumlah data yang sangat besar serta keberagaman gaya bahasa membuat proses analisis manual menjadi tidak efisien. Oleh karena itu, diperlukan pendekatan analisis sentimen berbasis komputasi untuk mengidentifikasi opini positif, negatif, dan netral dari ulasan-ulasan tersebut secara otomatis.

Analisis sentimen merupakan cabang dari Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi serta mengklasifikasikan opini publik terhadap suatu objek, produk, atau layanan [2]. Pendekatan ini dapat dilakukan dengan dua metode utama, yaitu *machine learning* dan *lexicon-based*. Metode *machine learning* membutuhkan data berlabel dalam jumlah besar agar model dapat belajar dari pola yang ada, sedangkan metode *lexicon-based* menggunakan kamus kata yang telah diberikan skor sentimen positif atau negatif untuk menentukan polaritas suatu teks [2].

Dalam penelitian ini, pendekatan *lexicon-based* digunakan untuk melakukan pelabelan awal terhadap data ulasan pengguna aplikasi Ruangguru. Setelah proses pelabelan, data direpresentasikan menggunakan metode Term Frequency–Inverse

Document Frequency (TF-IDF) untuk mengubah teks menjadi bentuk numerik yang dapat diproses oleh algoritma pembelajaran mesin [3]. Namun, hasil pelabelan awal menunjukkan adanya ketidakseimbangan jumlah data antar kelas sentimen, di mana ulasan negatif jauh lebih sedikit dibandingkan ulasan positif dan netral. Kondisi ini berpotensi menyebabkan bias dalam proses pembelajaran model.

Untuk mengatasi masalah tersebut, penelitian ini menerapkan teknik Synthetic Minority Over-sampling Technique (SMOTE) guna menyeimbangkan jumlah data pada setiap kelas sentimen [3]. Selain itu, algoritma Extreme Gradient Boosting (XGBoost) digunakan sebagai model klasifikasi utama karena terbukti efektif dalam menangani data berukuran besar dan kompleks [6]. Agar model bekerja secara optimal, dilakukan pula proses hyperparameter tuning menggunakan metode *GridSearchCV* yang telah terbukti efektif dalam menemukan kombinasi parameter terbaik untuk meningkatkan kinerja model [5].

Dengan demikian, penelitian ini difokuskan untuk menganalisis pengaruh penerapan *hyperparameter tuning* terhadap peningkatan akurasi model XGBoost dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Ruangguru. Pendekatan ini diharapkan mampu menunjukkan sejauh mana konfigurasi parameter berperan dalam meningkatkan kemampuan generalisasi model, sekaligus memperkuat efektivitas metode *Lexicon-Based*, TF-IDF, dan SMOTE sebagai tahapan pendukung dalam menghasilkan model klasifikasi yang optimal.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini disusun untuk memberikan arah yang jelas terhadap fokus penelitian serta metode yang digunakan. Berdasarkan latar belakang yang telah dijelaskan, penelitian ini mengidentifikasi permasalahan, metode yang diusulkan, serta pertanyaan penelitian sebagai berikut :

1.2.1 Permasalahan

Aplikasi Ruangguru sebagai salah satu platform pembelajaran daring terbesar di Indonesia telah banyak digunakan oleh pelajar dan masyarakat umum. Setiap pengguna memiliki pengalaman yang berbeda dalam menggunakan aplikasi tersebut dan sering kali menyampaikan pendapatnya melalui berbagai platform digital. Namun, jumlah ulasan yang sangat banyak dan bersifat tidak terstruktur menyebabkan sulitnya melakukan analisis secara manual.

Selain itu, hasil pelabelan sentimen berdasarkan pendekatan *Lexicon-Based* sering kali menunjukkan ketidakseimbangan antar kelas (positif, negatif, dan netral), yang dapat memengaruhi performa model klasifikasi. Kondisi ini menimbulkan kebutuhan akan pendekatan komputasi yang mampu mengelola data tidak seimbang sekaligus meningkatkan akurasi model analisis sentimen.

1.2.2 Metode Usulan

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan kombinasi beberapa teknik dan algoritma yang telah terbukti efektif berdasarkan kajian literatur sebelumnya, antara lain:

1. **Pelabelan Sentimen** menggunakan pendekatan *Lexicon-Based* berbasis kamus sentimen Bahasa Indonesia.
2. **Ekstraksi Fitur** menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)* untuk mengubah teks menjadi representasi numerik.
3. **Penyeimbangan Data** dengan metode *Synthetic Minority Over-sampling Technique (SMOTE)* agar distribusi kelas menjadi seimbang.
4. **Klasifikasi Sentimen** menggunakan algoritma *Extreme Gradient Boosting (XGBoost)* yang mampu menangani data berukuran besar dan kompleks.
5. **Optimasi Model** dengan menerapkan *Hyperparameter Tuning* melalui *GridSearchCV* untuk memperoleh konfigurasi parameter terbaik.

Pendekatan ini diharapkan mampu menghasilkan model analisis sentimen yang akurat, seimbang, serta mampu menangkap pola opini pengguna terhadap aplikasi Ruangguru secara komprehensif.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan dan metode yang diusulkan, maka pertanyaan penelitian yang ingin dijawab melalui penelitian ini adalah sebagai berikut:

1. Bagaimana tahapan proses analisis sentimen terhadap ulasan pengguna aplikasi Ruangguru dengan menggunakan metode *Lexicon-Based*, *TF-IDF*, *SMOTE*, dan *XGBoost*?
2. Bagaimana pengaruh penerapan *Hyperparameter Tuning* terhadap peningkatan performa algoritma *XGBoost* dalam klasifikasi sentimen?

3. Bagaimana kombinasi metode *Lexicon-Based*, *TF-IDF*, *SMOTE*, dan *XGBoost* dengan optimasi parameter dapat menghasilkan model analisis sentimen yang akurat dan seimbang?

1.3 Tujuan dan Manfaat

Penelitian ini memiliki tujuan dan manfaat yang dirancang untuk menjawab rumusan masalah yang telah dijelaskan sebelumnya. Tujuan penelitian menjelaskan sasaran yang ingin dicapai, sedangkan manfaat penelitian menggambarkan kontribusi yang diharapkan dari hasil penelitian ini.

1.3.1 Tujuan

Secara umum, penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi Ruangguru dengan memanfaatkan pendekatan *Lexicon-Based* dan algoritma *Extreme Gradient Boosting (XGBoost)* yang dioptimasi melalui *Hyperparameter Tuning*. Pendekatan ini diharapkan dapat menghasilkan model klasifikasi yang lebih akurat dan seimbang dalam mengenali opini pengguna terhadap layanan aplikasi.

Secara khusus, tujuan penelitian ini adalah:

1. Mendeskripsikan tahapan proses analisis sentimen terhadap ulasan pengguna aplikasi Ruangguru, meliputi pengumpulan data, *text preprocessing*, pelabelan menggunakan *Lexicon-Based*, ekstraksi fitur menggunakan *TF-IDF*, penyeimbangan data dengan *SMOTE*, serta klasifikasi menggunakan algoritma *XGBoost*.
2. Menganalisis pengaruh penerapan *Hyperparameter Tuning* terhadap peningkatan performa algoritma *XGBoost* dalam mengklasifikasikan sentimen positif, negatif, dan netral.
3. Menghasilkan model analisis sentimen yang akurat, seimbang, dan dapat merepresentasikan persepsi pengguna aplikasi Ruangguru secara objektif berdasarkan ulasan berbahasa Indonesia.

1.3.2 Manfaat

Hasil penelitian ini diharapkan dapat memberikan manfaat dari berbagai aspek, baik akademik, praktis, maupun teknis, sebagai berikut:

1. Manfaat Akademik

Penelitian ini dapat menjadi referensi dan acuan bagi mahasiswa atau peneliti lain yang ingin melakukan penelitian dalam bidang analisis sentimen berbasis *machine learning* dan *lexicon-based* pada teks berbahasa Indonesia.

2. Manfaat Praktis

Hasil penelitian ini dapat dimanfaatkan oleh pengembang aplikasi Ruangguru sebagai bahan evaluasi dalam meningkatkan kualitas layanan berdasarkan persepsi pengguna yang diperoleh secara otomatis melalui analisis sentimen.

3. Manfaat Teknis

Penelitian ini menunjukkan efektivitas penerapan *Hyperparameter Tuning* pada algoritma XGBoost dalam meningkatkan performa klasifikasi. Temuan ini dapat digunakan sebagai dasar pengembangan model serupa pada penelitian dengan karakteristik data tidak seimbang.

1.4 Batasan Masalah

Agar penelitian berjalan secara terarah dan fokus, maka batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan berupa ulasan pengguna aplikasi Ruangguru yang diperoleh dari lima platform digital: Google Play Store, Twitter (X), Instagram, TikTok, dan YouTube.
2. Proses pelabelan sentimen dilakukan menggunakan pendekatan *Lexicon-Based* berbasis kamus sentimen Bahasa Indonesia (INSET dan AFINN).
3. Ekstraksi fitur dilakukan dengan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*.
4. Penyeimbangan data dilakukan menggunakan metode *Synthetic Minority Over-sampling Technique (SMOTE)*.
5. Algoritma klasifikasi yang digunakan adalah *Extreme Gradient Boosting (XGBoost)*.
6. Optimasi model dilakukan menggunakan *GridSearchCV* untuk menentukan kombinasi parameter terbaik yang memberikan hasil akurasi tertinggi.

1.5 Sistematika Penulisan

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

1 BAB I PENDAHULUAN

Berisi latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, dan sistematika penulisan.

2 BAB II KAJIAN PUSTAKA

Menjelaskan teori-teori yang mendukung penelitian, seperti analisis sentimen, metode *Lexicon-Based*, TF-IDF, SMOTE, XGBoost, *hyperparameter tuning*, serta penelitian terdahulu yang relevan.

3 BAB III METODOLOGI PENELITIAN

Menguraikan tahapan penelitian yang dilakukan, mulai dari pengumpulan data, prapemrosesan teks, pelabelan, ekstraksi fitur, penyeimbangan data, klasifikasi, optimasi parameter, hingga evaluasi hasil.

4 BAB IV HASIL DAN PEMBAHASAN

Menyajikan hasil implementasi dan pembahasan terkait performa model serta analisis hasil klasifikasi sentimen.

5 BAB V PENUTUP

Berisi kesimpulan dan saran berdasarkan hasil penelitian yang telah dilakukan.