

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah membawa perubahan besar dalam dunia pendidikan, salah satunya melalui hadirnya platform pembelajaran daring. Di Indonesia, salah satu aplikasi yang paling populer adalah Ruangguru, yang menawarkan layanan belajar interaktif bagi siswa dari berbagai jenjang pendidikan. Dengan tingginya jumlah pengguna, Ruangguru menerima banyak ulasan di berbagai platform digital seperti Google Play Store, YouTube, TikTok, Twitter, dan Instagram.

Ulasan-ulasan tersebut mencerminkan persepsi dan pengalaman pengguna, baik positif maupun negatif, terhadap kualitas layanan, fitur aplikasi, dan efektivitas pembelajaran yang disediakan. Oleh karena itu, penting untuk menganalisis ulasan pengguna agar pengembang dapat memahami kebutuhan serta tingkat kepuasan pengguna secara objektif.

Salah satu pendekatan yang banyak digunakan dalam menganalisis opini publik adalah analisis sentimen, yaitu proses mengidentifikasi polaritas emosi (positif, negatif, atau netral) dari teks [1]. Dalam penelitian ini, pendekatan lexicon-based digunakan untuk melakukan pelabelan sentimen, yang kemudian dikombinasikan dengan metode TF-IDF (Term Frequency–Inverse Document Frequency) sebagai teknik ekstraksi fitur dan algoritma AdaBoost untuk klasifikasi [2][3].

Namun, penelitian sejenis kerap menghadapi masalah ketidakseimbangan data (imbalanced dataset), di mana jumlah data pada satu kelas (misalnya positif) jauh lebih banyak dibanding kelas lainnya [4]. Hal ini dapat menyebabkan model bias dan menurunkan performa klasifikasi. Untuk mengatasi hal tersebut, digunakan metode SMOTE (Synthetic Minority Oversampling Technique) yang menyeimbangkan distribusi data, serta GridSearchCV untuk melakukan penyesuaian parameter (hyperparameter tuning) sehingga kinerja model dapat dioptimalkan [5][6].

Hasil dari penerapan metode gabungan ini menunjukkan peningkatan akurasi dari 75,8% menjadi 84,23%, dengan F1-score rata-rata sebesar 0,84, yang berarti model mampu memberikan hasil klasifikasi yang lebih baik dan seimbang [7].

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana melakukan analisis sentimen terhadap ulasan pengguna aplikasi Ruangguru secara efektif menggunakan pendekatan lexicon-based?
2. Bagaimana penerapan metode TF-IDF dan AdaBoost dalam klasifikasi sentimen ulasan pengguna?
3. Bagaimana pengaruh penerapan SMOTE dan GridSearchCV terhadap peningkatan performa model klasifikasi sentimen?

1.2.1 Permasalahan

Permasalahan utama dalam penelitian ini meliputi:

1. Ketidakseimbangan jumlah data antara kelas positif dan negatif yang menyebabkan bias pada hasil klasifikasi.
2. Kompleksitas bahasa dalam ulasan pengguna yang dapat memengaruhi akurasi analisis sentimen.
3. Perlunya optimasi parameter model agar menghasilkan kinerja terbaik dalam klasifikasi sentimen.

1.2.2 Metode Usulan

Metode yang diusulkan dalam penelitian ini terdiri dari beberapa tahap, yaitu:

1. Pengumpulan Data: Mengambil data ulasan pengguna aplikasi Ruangguru dari platform digital seperti Google Play Store.
2. Preprocessing Data: Meliputi pembersihan teks, penghapusan tanda baca, stopwords removal, dan stemming.
3. Pelabelan Sentimen: Menggunakan pendekatan lexicon-based untuk menentukan polaritas setiap ulasan.
4. Ekstraksi Fitur: Menggunakan TF-IDF untuk mengubah teks menjadi representasi numerik.
5. Penyeimbangan Data: Menerapkan SMOTE untuk mengatasi ketidakseimbangan kelas.
6. Pelatihan Model: Menggunakan algoritma AdaBoost sebagai metode klasifikasi, dengan GridSearchCV untuk optimasi parameter.
7. Evaluasi Model: Menggunakan metrik akurasi, precision, recall, dan F1-score untuk menilai performa model.

1.2.3 Pertanyaan Penelitian

1. Apakah pendekatan lexicon-based efektif dalam memberikan label sentimen terhadap ulasan pengguna Ruangguru?

2. Seberapa besar pengaruh metode TF-IDF terhadap hasil klasifikasi sentimen?
3. Bagaimana kontribusi SMOTE dalam memperbaiki keseimbangan data dan meningkatkan performa model?
4. Apakah penerapan GridSearchCV pada algoritma AdaBoost dapat meningkatkan akurasi klasifikasi sentimen secara signifikan?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Tujuan penelitian ini adalah:

1. Melakukan analisis sentimen terhadap ulasan pengguna aplikasi Ruangguru menggunakan pendekatan lexicon-based.
2. Mengimplementasikan metode TF-IDF dan AdaBoost dalam proses klasifikasi sentimen.
3. Mengatasi ketidakseimbangan data dengan metode SMOTE untuk meningkatkan akurasi model.
4. Mengoptimalkan performa model melalui tuning parameter menggunakan GridSearchCV.
5. Mengevaluasi hasil klasifikasi dengan berbagai metrik kinerja untuk mengetahui efektivitas metode yang digunakan.

1.3.2 Manfaat

Manfaat yang diharapkan dari penelitian ini adalah:

1. Bagi Pengembang Aplikasi: Memberikan masukan terkait aspek yang perlu diperbaiki berdasarkan opini pengguna.
2. Bagi Peneliti: Menjadi referensi dalam penerapan metode pembelajaran mesin untuk analisis sentimen pada data tidak seimbang.
3. Bagi Akademisi: Memberikan kontribusi terhadap pengembangan penelitian di bidang text mining dan analisis sentimen berbasis machine learning.

1.4 Batasan Masalah

Batasan penelitian ini adalah sebagai berikut:

1. Data ulasan yang digunakan berasal dari beberapa platform yaitu Google Play Store, Youtube, Tiktok, Twitter, dan Instagram.
2. Pelabelan sentimen dilakukan secara otomatis menggunakan pendekatan lexicon-based.
3. Metode ekstraksi fitur dibatasi pada TF-IDF, dan algoritma klasifikasi menggunakan AdaBoost.
4. Proses penyeimbangan data hanya menggunakan SMOTE.

5. Evaluasi kinerja model dilakukan dengan metrik akurasi, precision, recall, dan F1-score.

1.5 Sistematika Penulisan

Sistematika penulisan penelitian ini dibagi dalam 3 bab yaitu:

1. BAB I PENDAHULUAN

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

2. BAB II KAJIAN PUSTAKA

Berisi teori-teori yang mendukung penelitian, seperti analisis sentimen, lexicon-based approach, TF-IDF, AdaBoost, SMOTE, dan penelitian terdahulu.

3. BAB III METODOLOGI PENELITIAN

Menjelaskan langkah-langkah penelitian mulai dari pengumpulan data hingga evaluasi model.

4. BAB IV HASIL DAN PEMBAHASAN

Menyajikan hasil eksperimen, analisis performa model, serta pembahasan terkait hasil penelitian.

5. BAB V PENUTUP

Menyimpulkan hasil penelitian dan memberikan rekomendasi untuk pengembangan penelitian selanjutnya.