

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

YouTube bukan hanya platform untuk berbagi dan menonton video, tetapi juga ruang interaksi sosial yang memungkinkan pengguna untuk meninggalkan komentar dan berkomunikasi dengan sesama pengguna. Komentar-komentar ini menjadi wadah ekspresi opini, mencerminkan respons dan persepsi pengguna terhadap konten yang ditonton [1]. Klasifikasi terhadap komentar-komentar tersebut sangat penting untuk memahami pola sentimen, membantu kreator konten, pelaku bisnis, dan peneliti dalam menganalisis tren opini publik.

Mengelola dan mengklasifikasi komentar *YouTube* menjadi tantangan karena volumenya besar dan mengandung sentimen berharga [2]. Tanpa teknik analisis yang tepat, wawasan ini bisa terabaikan, menyulitkan pemahaman opini publik. Kompleksitas bahasa, seperti slang, emotikon, dan konteks budaya, makin memperumit klasifikasi sentimen. Oleh karena itu, penelitian ini krusial untuk mengembangkan metode klasifikasi yang lebih efisien dan dapat diterapkan dalam pemantauan opini publik dan analisis tren media sosial.

Untuk mengatasi tantangan tersebut, penelitian ini menerapkan *Grid Search* sebagai metode optimasi untuk meningkatkan akurasi klasifikasi. *Grid Search* memungkinkan pencarian kombinasi parameter terbaik untuk model machine learning, sehingga dapat meningkatkan hasil prediksi [3]. Di sisi lain, teknik ekstraksi fitur seperti *TF-IDF* (*Term Frequency-Inverse Document Frequency*) juga berperan penting dalam meningkatkan kualitas klasifikasi sentimen dengan menyoroti kata-kata yang paling relevan dalam teks [4].

Algoritma seperti *SVM* (*Support Vector Machine*), *Logistic Regression*, dan *Random Forest* terbukti efektif digunakan dalam klasifikasi sentimen [5], [6], [7]. Penelitian ini mengoptimalkan performa ketiga algoritma tersebut menggunakan *Grid Search* pada komentar *YouTube*. Selain itu, penelitian ini membandingkan performa model sebelum dan sesudah optimasi untuk menilai efektivitasnya.

Diharapkan hasilnya dapat meningkatkan akurasi analisis sentimen otomatis serta mendukung pemanfaatan data komentar *YouTube* secara optimal.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang masalah, rumusan masalah pada penelitian ini adalah bagaimana kinerja algoritma klasifikasi *SVM*, *Logistic Regression*, dan *Random Forest* yang dioptimasi menggunakan *Grid Search* dalam klasifikasi sentimen komentar *YouTube*?

1.3 Batasan Masalah

Dalam penelitian ini, terdapat beberapa batasan yang perlu dijelaskan untuk memastikan fokus yang jelas pada ruang lingkup yang diteliti. Batasan-batasan tersebut antara lain sebagai berikut.

1. Data yang digunakan pada penelitian ini adalah teks komentar yang diambil dari *YouTube*. Fokus penelitian ini hanya pada komentar-komentar yang terdapat di platform *YouTube*, tanpa memperhitungkan jenis atau kategori video yang diunggah.
2. Tahap *preprocessing* yang dilakukan yaitu *case folding*, *text cleaning*, *slang words Normalization* menggunakan kamus dari *github*, *stopwords removal* dengan *corpus stopwords* bahasa Indonesia dari *library nltk*, *tokenizing*, dan *stemming* dengan *library sastrawi*.
3. *Labeling* menggunakan kamus sentimen dari *github*.
4. Kategori yang dipakai untuk klasifikasi sentimen yaitu positif, negatif, dan netral.
5. Ekstraksi fitur dari teks komentar menggunakan *TF-IDF*.
6. *Oversampling* menggunakan *SMOTE*.
7. Algoritma klasifikasi yang dipakai yaitu *SVM*, *Logistic Regression*, dan *Random Forest*.
8. Optimasi yang dilakukan menggunakan metode *Grid Search*.

1.4 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan adalah:

1. Mengimplementasikan metode klasifikasi *Support Vector Machine (SVM)*, *Logistic Regression*, dan *Random Forest* menggunakan *Grid Search* untuk meningkatkan akurasi dan performa model klasifikasi sentimen komentar YouTube.
2. Mengevaluasi performa dari metode *Support Vector Machine (SVM)*, *Logistic Regression*, dan *Random Forest* sebelum dan sesudah dilakukan optimasi.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang *machine learning* dan klasifikasi sentimen. Penerapan algoritma *Support Vector Machine (SVM)*, *Logistic Regression*, dan *Random Forest* dengan optimasi *Grid Search* serta teknik ekstraksi fitur *TF-IDF* dapat menjadi referensi bagi penelitian sejenis dalam meningkatkan akurasi klasifikasi sentimen teks, khususnya pada komentar YouTube.

Penelitian ini juga dapat digunakan sebagai acuan bagi pengembang teknologi dan praktisi untuk mengimplementasikan model klasifikasi sentimen yang dihasilkan. Selain itu, hasil penelitian ini juga dapat dimanfaatkan oleh perusahaan penyedia jasa digital, seperti platform media sosial, agensi pemasaran, atau penyedia layanan analitik, untuk memahami opini pengguna secara lebih akurat. Dengan model klasifikasi sentimen yang lebih optimal, perusahaan dapat meningkatkan layanan mereka, menyusun strategi pemasaran berbasis data, serta mengidentifikasi tren opini publik dengan lebih efektif. Hasil penelitian ini juga dapat menjadi dasar bagi penelitian lanjutan yang berfokus pada optimasi model dan penerapan *machine learning* untuk data teks.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun ke dalam lima bab, berikut uraian secara garis besar dari masing-masing bab:

BAB I PENDAHULUAN, menjelaskan latar belakang permasalahan yang mendasari penelitian, rumusan masalah yang diidentifikasi, batasan yang digunakan untuk membatasi ruang lingkup penelitian, tujuan yang ingin dicapai, manfaat dari penelitian, serta sistematika penulisan laporan penelitian.

BAB II TINJAUAN PUSTAKA, menguraikan kajian literatur yang relevan dengan topik penelitian, mencakup konsep klasifikasi sentimen, metode ekstraksi fitur menggunakan *TF-IDF*, metode klasifikasi seperti *SVM*, *Logistic Regression* dan *Random Forest*, serta teknik optimasi *Grid Search*. Selain itu, juga dibahas penelitian terdahulu yang relevan dengan penelitian ini.

BAB III METODE PENELITIAN, menjelaskan secara rinci tahapan-tahapan yang dilakukan dalam penelitian. Tahapan tersebut meliputi proses pengumpulan data, tahap preprocessing, penerapan metode ekstraksi fitur menggunakan *TF-IDF*, implementasi metode klasifikasi *SVM*, *Logistic Regression* dan *Random Forest*, optimasi parameter menggunakan *Grid Search*, dan evaluasi terhadap kinerja metode yang digunakan.

BAB IV HASIL DAN PEMBAHASAN, bagian ini memaparkan hasil-hasil yang diperoleh dari eksperimen yang telah dilakukan. Hasil tersebut mencakup perbandingan kinerja antara metode yang telah dioptimasi dan metode tanpa optimasi, serta analisis terhadap performa metode klasifikasi yang digunakan setelah optimasi.

BAB V PENUTUP, menyajikan kesimpulan dari hasil penelitian yang telah dilakukan, serta memberikan saran untuk pengembangan dan penelitian lanjutan yang dapat dilakukan di masa depan.