

Analisis Sentimen terhadap Ulasan Film Berbahasa Inggris Menggunakan Metode *Support Vector Machine* dengan Feature Selection *Information Gain*

Nauffan Muti Hibattullah¹, Said Al Faraby²,
Mahendra Dwifebri Purbolaksono³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹nauffanmufti@students.telkomuniversity.ac.id,

²saidalfaraby@telkomuniversity.ac.id,³mahendradp@telkomuniversity.ac.id

Abstrak

Analisis sentimen adalah suatu bidang yang menganalisis opini, sikap, dan emosi dari banyak orang terhadap suatu produk, jasa atau entitas lain, bidang penelitian ini cukup populer. Teks ulasan merupakan sebuah teks yang isinya berupa ulasan, *review* atau penilaian pada suatu karya seperti film, drama, dan juga buku. Teks ulasan memiliki fungsi yaitu untuk menilai, menimbang, serta mengajukan kritik pada karya maupun peristiwa yang diulas. Bahasa Inggris adalah bahasa yang paling banyak digunakan di dunia. Ada 400 juta penutur asli dan 2 milyar orang mempelajarinya sebagai bahasa kedua. *Support Vector Machine* (SVM) merupakan salah satu metode dalam *supervised learning* yang biasanya digunakan untuk klasifikasi (*Support Vector Classification*) dan regresi (*Support Vector Regression*). Metode *Information Gain* adalah metode yang menggunakan teknik *scoring* untuk pembobotan sebuah fitur dengan menggunakan maksimal *entropy*. Fitur yang dipilih adalah fitur dengan nilai *Information Gain* yang lebih besar atau sama dengan nilai *threshold* tertentu. Penelitian ini menunjukkan proses dengan dilakukan kombinasi *Stopword* dan *Stemming* maka akurasi yang dihasilkan akan lebih maksimal, karena proses *preprocessing* lebih lengkap sebesar 86,12%. Selain itu didapatkan bahwa seleksi fitur *Information Gain* (IG) pada penelitian ini membuat akurasi menjadi rendah, tetapi dapat menjadi sebuah solusi yang cukup baik untuk mengatasi masalah *overfitting* pada pengujian analisis sentimen ini. Dan klasifikasi analisis sentimen *movie review* sangat tepat menggunakan algoritma *Support Vector Machine* (SVM) kernel *Linear*. Dikarenakan *Linear* berfokus dengan fitur-fitur yang mengandung nilai biner yang didapatkan dari *hyperplane* terbaik dengan memaksimalkan jarak antar kelas.

Kata kunci: Analisis Sentimen, Ulasan, Film, Berbahasa Inggris, *Support Vector Machine*, *Information Gain*
