

ABSTRAK

Film merupakan bagian dari gaya hidup digital terkini. Sebagai konsekuensi, banyak aplikasi *video streaming* bermunculan. Ulasan film dapat dijadikan alternatif untuk memilih film yang akan ditonton selanjutnya. Namun, ulasan mengenai suatu film sangat mungkin berisikan bocoran mengenai konten film yang tidak diinginkan penggemar film. Sebenarnya, hal yang mereka hanyalah informasi sentimen yang ada pada ulasan film tersebut. Analisis sentimen dapat dijadikan solusi untuk mendapatkan informasi sentimen pada ulasan film. Analisis sentimen, juga dikenal sebagai kategorisasi sentimen, adalah studi yang menganalisa informasi subjektif pada suatu objek tertentu.

Sayangnya, representasi *bag of words* yang digunakan dalam analisis sentimen bermasalah dalam penanganan matriks fitur berdimensi tinggi. Permasalahan ini bisa ditangani dengan memilih fitur menggunakan skema seleksi fitur tertentu. Fitur yang bagus adalah fitur yang memiliki relevansi tinggi dengan kelas output. Studi terkini dalam seleksi fitur untuk analisis sentimen berhasil menyaring fitur yang sangat relevan namun fitur-fitur tersebut sangat jarang muncul. Akibatnya, analisis sentimen menghadapi permasalahan *over-fitting*.

Information Gain adalah metode yang lazim digunakan untuk memilih fitur. Penelitian ini mengajukan suatu skema seleksi fitur dan klasifikasi berdasarkan nilai *Information Gain*. Skema seleksi fitur yang diusulkan, IGDFFS, memilih fitur yang setidaknya-tidaknya memenuhi dua kriteria berikut: (1) relevan dengan kelas output dan (2) berpeluang besar untuk muncul. Berdasarkan nilai *Information Gain* juga, polaritas kamus dibangun. Skema klasifikasi yang diusulkan, IGC, menggunakan kamus tersebut untuk mengklasifikasikan ulasan film. Hasil percobaan menunjukkan bahwa kombinasi IGDFFS dan IGC, yang mencapai akurasi sebesar 96%, dapat dikatakan lebih efektif daripada metode yang diajukan pada penelitian sebelumnya.

Kata kunci: Analisis Sentimen, Seleksi Fitur, Klasifikasi, Information Gain