

Abstrak

Terdapat banyak film yang dapat menjadi pilihan untuk ditonton, namun belum tentu film-film tersebut memiliki kualitas yang bagus. Ulasan film menjadi sangat penting dikarenakan mampu memberikan informasi mengenai kualitas suatu film sehingga dapat dijadikan referensi untuk menonton suatu film. Dikarenakan data mengenai ulasan film berjumlah sangat banyak dan membutuhkan waktu lama jika membacanya secara manual satu per satu, diperlukan sistem yang dapat melakukan analisis sentimen dalam waktu yang singkat. Metode *machine learning* seperti *K-Nearest Neighbor* (KNN) dapat digunakan untuk melakukan analisis sentimen, namun memiliki performansi yang kurang baik jika fitur yang digunakan tidak ditentukan dengan baik. Untuk menentukan fitur yang akan digunakan, penggunaan ekstraksi fitur dan seleksi fitur perlu digunakan. Dalam penelitian ini, ekstraksi fitur yang digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF) dan seleksi fitur yang digunakan adalah *chi-square*. Dengan kombinasi antara metode *machine learning* KNN, ekstraksi fitur TF-IDF, dan seleksi fitur *chi-square*, dapat memberikan hasil evaluasi performansi terbaik pada $K = 37$ dengan nilai akurasi akurasi 80.75%, *precision* 83.24%, *recall* 77.00%, dan *F1-measure* 80.00%

Kata kunci : analisis sentimen, TF-IDF, *chi-square*, KNN