

Abstrak

Lebih dari 80% informasi pada dunia maya disimpan dalam bentuk teks, sehingga penambangan teks memiliki potensi yang lebih tinggi dibandingkan dengan penambangan data. *Sentiment analysis* adalah cabang baru pada penambangan teks, bertujuan untuk mendeteksi polaritas pada sebuah teks, kalimat, paragraf, maupun dokumen. *Sentiment analysis merupakan* baik untuk menganalisis *movie review* terkait ulasan yang jelas mengekspresikan opini pada sebuah film, seperti positif atau negatif. Penelitian ini bertujuan untuk menganalisis performansi sentimen pada Internet Movie Database yang berisi 50.000 ulasan film menggunakan metode *feature selection* Information Gain dan Decision Tree Classifier. Information Gain dapat mengurangi fitur yang tidak relevan dan Decision Tree memiliki akurasi yang baik pada banyak penelitian terkait. Untuk meningkatkan performansi dan hasil akurasi digunakan *feature extraction* dan model N-gram untuk mendapatkan fitur dengan struktur gramatikal yang baik pada *movie review*. Hasil pengujian dibagi dengan parameter N-gram dan hasil terbaiknya akan digunakan pada perbandingan antara metode DT dengan DT(Pruning), SVM, MNB, dan RF menggunakan IG *Threshold*. Hasil akurasi terbaik yang didapatkan klasifikasi DT adalah 72.69% dengan parameter kombinasi Unigram, Bigram dan Trigram dan IG *Threshold* 0.1.

Kata kunci : sentiment analysis, movie review, information gain, decision tree, N-gram, threshold