

ABSTRAK

Perkembangan film di Indonesia akhir-akhir ini semakin meningkat, mulai dari segi alur cerita, isi film dan aktor aktrisnya. Pada penelitian ini penulis ingin memprediksi jumlah penonton film Indonesia di twitter. Twitter adalah sebuah media social yang digunakan untuk mencari data komentar penonton melalui *tweet-tweet* pengguna twitter. Melalui sebuah prediksi diharapkan production house dan produser menyusun strategi agar film yang akan ditayangkan laku dipasaran. Pada penelitian sebelumnya dilakukan menggunakan data film *Hollywood*, akan tetapi pada penelitian ini menggunakan data film Indonesia. Pada penelitian ini melakukan dua pendekatan melalui *user* dan *tweet*. Fitur-fitur yang pendekatan menggunakan data *user*, yaitu *follower official*, *follower production house*, *follower produser*, *follower sutradara*, dan *follower 3 artis pemeran utama dalam film*, sedangkan fitur-fitur dari data *tweet*, yaitu *retweet*, *like*, jumlah *tweet* yang mengandung URL. Tujuan dari penelitian ini adalah untuk menerapkan proses *pre-processing* untuk membentuk data masukan yang lebih cocok dimasukkan kedalam proses algoritma *naïve bayes classifier*. Hasil pengujian menggunakan algoritma *naïve bayes classifier* adalah sebesar 78.79%. Untuk meningkatkan performansi dan mencegah overfitting dari sistem yang dibangun, dapat dilakukan dengan menambahkan *feature subset selection* pada *naïve bayes classifier* yang mempunyai rata-rata akurasi sebesar 77.96%.

Kata kunci: Twitter, *Crawling Data*, *Pre-processing*, *Naive Bayes Classifier*, *Measuring performance*, Film.