

ABSTRAK

Banyaknya opini – opini yang ada dalam media sosial Twitter salah satunya mengenai film produksi Indonesia seringkali menyulitkan masyarakat untuk menarik kesimpulan terhadap sebuah film, karena banyaknya opini masyarakat yg berbeda mengenai film yang diproduksi oleh sutradara Indonesia. Hal tersebut juga seringkali dimanfaatkan oleh para pengguna untuk mempromosikan produk dagangan mereka melalui hastag yang berkaitan dengan film Indonesia. Promosi produk yang banyak dilakukan sering kali mengganggu pengguna karena konten yang tidak diperlukan namun secara terus menerus muncul dan memberikan dampak negatif ke pengguna secara umum dapat dikategorikan sebagai spam. Maka dari itu perlu adanya analisis sentimen pada twitter terhadap film khususnya film produksi Indonesia dengan menerapkan sebuah *filter* untuk mengidentifikasi *tweet* tersebut sebuah *spam* atau bukan spam, agar tidak memperburuk hasil klasifikasi sentimen tersebut.

Salah satu algoritma klasifikasi teks yang cukup bagus adalah algoritma *Convolutional Neural Network* (CNN) yang menunjukkan bahwa CNN mempunyai kelebihan yaitu melakukan pengenalan sentimen pada teks dengan memperhitungkan *sequence* kata pada kalimat, tidak seperti metode classical yang tidak memperdulikan urutan kata. Penelitian ini menggunakan metode CNN arsitektur *multi-filter* untuk mengurangi proses *tuning parameter* dan menggunakan *spam filtering* untuk membuang ulasan berupa *spam*. Penelitian ini menggunakan pengujian *confusion matrix* dan menggunakan pendekatan *k-fold cross validation* dengan perbandingan 80% untuk data latih dan 20% untuk data uji.

Berdasarkan implementasi dan hasil pengujian yang telah dilakukan didapatkan hasil bahwa model CNN yang telah dibuat dengan arsitektur *multi-kernel* dengan ukuran *kernel* sebesar (3,4,5) dengan jumlah *filter* 256 di setiap ukuran *filter* mendapatkan hasil rata-rata akurasi sebesar 71,77%, presisi sebesar 74,24% dan *recall* sebesar 70,07%. Ini membuktikan bahwa model CNN yang dibuat dengan *multi-kernel* lebih baik jika dibandingkan dengan model CNN yang dibuat tanpa arsitektur *multi-kernel* yang mendapatkan hasil rata-rata akurasi 65,05%, presisi 64,55%, *recall* 62,91%. Kemudian dilakukan proses *spam filtering* sebelum dilakukan proses klasifikasi sentimen dengan model CNN dan mendapatkan peningkatan hasil menjadi rata-rata akurasi sebesar 75,40%, presisi sebesar 74,84% dan *recall* sebesar 71,97%. Hal ini membuktikan bahwa penggunaan *spam filtering* sebelum dilakukan proses klasifikasi sentimen berhasil meningkatkan akurasi sebesar 3.78%.

Kata Kunci: ulasan film twitter, analisis sentimen, CNN