

ABSTRAK

Di Indonesia menonton film yang terdapat bahasa kasar pada subtitle merupakan hal yang mengganggu. Namun perbedaan budaya mengakibatkan beberapa film dari luar negeri menganggap bahasa kasar sebagai hal yang biasa. Klasifikasi teks bahasa kasar merupakan salah satu hal yang dapat dilakukan untuk menangani hal tersebut. Penelitian ini menggunakan metode *supervised learning* berupa *support vector machine* (SVM) untuk melakukan klasifikasi teks bahasa kasar. Teks pada subtitle diklasifikasikan kedalam 2 kategori yaitu teks kategori bahasa dan kategori teks bukan bahasa kasar.

SVM merupakan metode *supervised learning* yang memiliki kelemahan dalam menentukan parameter paling optimal. Sedangkan pemilihan parameter yang tepat pada SVM akan sangat berpengaruh terhadap hasil *accuracy* klasifikasi. Penelitian ini menggunakan metode optimasi parameter *grid search* (GS) untuk mendapatkan parameter SVM yang paling optimal. Metodologi penelitian ini dimulai dari studi literatur, persiapan dataset, *preprocessing*, pembobotan kata menggunakan *term frequency-inverse document frequency* (TF-IDF), pembuatan model tanpa optimasi, pembuatan model dengan optimasi, validasi dan pengujian menggunakan *k-fold cross validation*, kesimpulan dan saran.

Hasil pengujian model SVM tanpa optimasi parameter untuk melakukan klasifikasi teks bahasa kasar pada subtitle film didapatkan nilai *accuracy* sebesar 92.3%, nilai *precision* sebesar 99%, nilai *recall* sebesar 85.2%, nilai *f1-score* sebesar 91.5%. Sedangkan hasil dari pengujian model SVM model dengan optimasi parameter *grid search* untuk melakukan klasifikasi teks bahasa kasar pada subtitle film didapatkan nilai *accuracy* sebesar 95.4%, nilai *precision* sebesar 98.3%, nilai *recall* sebesar 92.2%, nilai *f1-score* sebesar 95.1%. Hasil pengujian tersebut menunjukkan penerapan algoritma *grid search* untuk membantu menentukan parameter SVM dapat meningkatkan *accuracy* sebesar 3.1% dari 92.3% menjadi 95.4%. Hal tersebut juga berlaku pada *recall* dengan peningkatan nilai sebesar 7.0% dari 85.2% menjadi 92.2% dan peningkatan *f1-score* sebesar 3.6% dari 91.5% menjadi 95.1%. Namun memberikan penurunan nilai pada *precision* sebesar 0.7% dari 99% menjadi 98.3%. Hal tersebut dapat disimpulkan bahwa algoritma optimasi parameter *grid search* merupakan algoritma pemilihan parameter yang mampu meningkatkan *accuracy* klasifikasi teks bahasa kasar pada metode SVM.

Kata kunci: SVM, *Grid search*, Klasifikasi Teks, Bahasa Kasar, Subtitle Film.