

ABSTRAK

Tayangan televisi membutuhkan rating dalam penilaiannya, namun dibutuhkan juga opini masyarakat untuk melengkapi penilaian tersebut. Sehingga dibutuhkan analisis sentimen dalam penyelesaiannya. Tahapan yang penting dalam analisis sentimen yaitu *preprocessing* karena dalam opini masyarakat masih banyak penulisan yang tidak sesuai. Tujuan dari penelitian ini yaitu untuk membandingkan hasil performa dengan menggunakan perbedaan skenario *preprocessing* untuk mendapatkan kinerja *preprocessing* terbaik pada *Support Vector Machine* dan *Naïve Bayes* terhadap analisis sentimen mengenai acara televisi X Factor Indonesia. Tahapan yang digunakan dimulai dari studi pustaka, analisis permasalahan, perancangan, pengumpulan data, *preprocessing* dengan dua skenario, pemmbobotan kata dengan TF-IDF, klasifikasi menggunakan *Support Vector Machine* dan *Naïve Bayes*, kemudian dihasilkan akurasi dari *Confusion Matrix*. Hasil penelitian ini menunjukkan bahwa kinerja terbaik diperoleh dengan menggunakan skenario *preprocessing* yang lebih lengkap, meliputi *casefolding*, *remove emoji*, *cleansing*, *remove repetition character*, normalisasi kata, *negation handling*, *stopwords removal*, *stemming* dan *tokenization* dengan akurasi sebesar 79,44% pada algoritma *Support Vector Machine*. Dengan adanya penelitian ini, maka dapat diketahui bahwa *preprocessing* yang lebih lengkap pada algoritma *Support Vector Machine* lebih baik dalam hal akurasi, presisi, *recall* dan *f-1 score*.

Kata kunci: Analisis sentimen, *Preprocessing*, *Support Vector Machine*, *Naïve Bayes*

ABSTRACT

Television shows need a rating in their assessment, but public opinion is also needed to complete the assessment. So that sentiment analysis is needed in its completion. An important step in sentiment analysis is preprocessing because in public opinion there are still many inappropriate writings. The purpose of this study is to compare the performance results using different preprocessing scenarios to get the best preprocessing performance on Support Vector Machine and Naïve Bayes on sentiment analysis about the television show X Factor Indonesia. The stages used start from literature study, problem analysis, design, data collection, preprocessing with two scenarios, word weighting with TF-IDF, classification using Support Vector Machine and Naïve Bayes, then resulting accuracy from Confusion Matrix. The results of this study indicate that the best performance is obtained using a more complete preprocessing scenario, including casefolding, remove emoji, cleansing, remove repetition character, word normalization, negation handling, stopwords removal, stemming and tokenization with an accuracy of 79.44% on the Support Vector Machine algorithm. With this research, it can be seen that the more complete preprocessing of the Support Vector Machine algorithm is better in terms of accuracy, precision, recall and f-1 score.

Keywords: Sentiment Analysis, Preprocessing; Support Vector Machines, Naïve Bayes