ABSTRAK

Song lyrics are one very important element in describing the expressions and emotions that exist in the song. The process of classifying songs can be done one of them by using song lyrics as a medium to classify the emotions contained in them. The level of accuracy of the song classification process is determined by the preprocessing process which consists of case folding, stop word filtering, tokenizing, and lemmatization. The lemmatization process itself is one of the important processes in the preprocessing stage where a word will be returned to its dictionary form so that it can reduce and clarify the features to carry out the classification process, so that the classification process can be carried out faster. In previous studies related to the Javanese language, there was no specific lemmatizer for Javanese, so in this study the lemmatizer was built by itself using the Lookup Dictionary method. The use of TF-IDF weighting and Support Vector Machine (SVM) carried out in this study with consideration, can compare the results of system performance between data through the lemmatization stage and not. The classification process begins with manual labeling, then data through the stages of preprocessing, feature extraction, weighting, training and classification using SVM. In the emotional classification process of Javanese songs using SVM, lemmatization, and TF-IDF, it can reduce the number of features as much as 5.89% where the data that goes through the lemmatization stage is 0.35% better than the data that not through the lemmatization stage.

Keywords: Javanese Language, Javanese Lemmatizer, Preprocessing, Text Classification