

## **ABSTRACT**

Lyrics are an important component to know the messages contained in a song. The lyrics can be used as a research object in text categorization. The current system for categorizing songs on application like iTunes on the iPhone, Laya on Android has been able to categorize songs by genre, artist, and album. However, there has been no categorization song by song topics, categorizing songs based on topic is still done manually. This research aims to create a system that categorizes songs by song topics using lyrics as research objects. This research begins by collecting lyrics from many website which serve as research dataset, then data preprocessing is done, consisting of case folding, tokenization, stopword removal and stemming. The next step is to do feature selection by using chi-square. Furthermore, feature extraction will be on the feature selection results by using bag of words. Classifier building is then performed to be used in the classification process. The next step is to classify song lyrics on test data into a class topic by using multinomial naïve bayes method. The last step in this research is to evaluate system by calculating accuracy, precision, recall and f1-measure. Topic class in this system are love, friendship, nasionalism, family, religion, and negative content. The performance of this system in this study, the f1-measure values at 88.91% and accuracy at 96.00%.

Keywords: classification, song lyrics, chi-square, multinomial naïve bayes