ABSTRACT

Film is one form of mass communication media that is used as a entertainment source for people. Film is effective on giving information. Nowdays, so many websites that provide information about the latest films. One of available information is about film genre. One film may have one or more genre. But to determine a film?s genre, people have to read the film?s synopsis first or to watch the film, it takes long time. This final project provides solution for the problem by classifying multigenre film from the film?s synopsis using naive bayesian multi-label classifier. There are several steps in this research such as undersampling technique, preprocessing, feature selection, classifier building, and classifier evaluation. On preprocessing step we use 3 methods: stop words removal, word segmentation and stemming, all of those method is applied on data that?s available on imdb.com/interfaces. On feature selection and classifier building process we perform chi square test and applying naive bayesian multi-label algorithm. Naive bayesian multi-label algorithm is a modified version of multinomial naive bayes that every class is combination of class which formed from data. The test result shows that the most effective method to classify film genre based on synopsis is to use Naive Bayesian Multi-Label (NBML) classifier using micro average f1-measure which gives result 87%, compared to other popular classifier (SVM classifier gives 65,15%, Decision Tree gives 50,53%, and K-NN gives 59,52%).

Keywords: *classification, multi-label, movie, classifier, naive bayesian*