

Abstract

Diabetes is a disease that cause an abnormally high levels of glucose in the blood and generalization of damage on blood vessels that affect the heart, eyes, kidneys, and nerves also cause various complications. Early identification of diabetes can be done using a model for prediction or diagnosis of diabetes. In this study, Random Forest was used to predict diabetics based on drug reviews content. Feature extraction is done by using N-gram and term frequency - inverse document frequency (TF-IDF). The best result of the experiments is accuracy of 95,2% with unigram feature extraction.

Key words: classification, random forest, N-gram, TF-IDF
