

Abstract

The existence of rumors on Twitter has caused a lot of unrest among Indonesians. Unrecognized validity confuses users for that information. In this study, an Indonesian rumor detection system is built by using J48 Algorithm in collaboration with Term Frequency Inverse Document Frequency (TF-IDF) weighting method. Dataset contains 47.449 tweets that have been manually labeled. This study offers new features, namely the number of emoticons in the display name, the number of digits in the display name, and the number of digits in the username. These three new features are used to maximize information about information sources. The highest accuracy is obtained by 75.55% using 90% training data and 1.000 TF-IDF features in 1-gram to 3-gram combinations.

Keywords: Twitter, rumor, pre-processing, J48, TF-IDF
