ABSTRACT

Twitter is one of the social media that is currently popularly used around the world. It's just that twitter has some problems that adversely affect its users. Hoax is one of the negative things that often occur in social media, news in the hoax is still doubt the truth or the fact. In this final project, the authors built a system to detect hoax news on twitter. The use of the Term Frequency Inverse Document Frequency (TF-IDF) weighting system in the system gives a weighted value to a tweet taken from the occurrence of a hoax news sent by someone on Twitter. Data classification uses the Support Vector Machine (SVM) method of the system to predict the possibility of a twitter account user spreading a hoax news based on the user's behavior. Testing data is done based on the contents of content tweets. Datasets are arranged based on attributes used such as the number of retweets, URLs, number of hashtags, provocations, feuds, anxieties, and unbalanced news. Processed data is divided into training data and testing data. The result of data using all features get the highest accuracy is 78,33%.

Keywords: TF-IDF, SVM, hoax, twitter, pre-processing, measuring performance