

ABSTRACT

The hot topic that is being discussed by the public has a variety of information. Twitter as social media has a system that can determine topics that are being discussed by users, called trending topics. However, Twitter has not been able to classify topics into several categories to get better information, the purpose of this study is to determine the effect of Chi-Square feature selection on the topic classification system into several categories of Indonesian tweet data taken from Twitter. Chi-Square is a feature selection technique that calculates the value of feature dependencies on a class. Selected features will be classified by the Naïve-Bayes Multinomial (NBM) method specifically designed for text classification. The test was conducted in several scenarios, and the scenario with the best results was obtained from a combination of 80: 20 dataset ratios, n-gram features in the form of unigram + bigram + trigram, and without feature selection with an accuracy value of 74.05%. It can be concluded that the Chi-Square feature selection does not have a significant effect on the trending topic prediction process.

Keywords: *twitter, trending topic, feature selection, chi-square, naïve-bayes multinomial, n-gram.*