

## Performance Comparison of ADTree and Naive Bayes Algorithms for Spam Filtering

**Authors :** Thanh Nguyen, Andrei Doncescu, Pierre Siegel

**Abstract :** Classification is an important data mining technique and could be used as data filtering in artificial intelligence. The broad application of classification for all kind of data leads to be used in nearly every field of our modern life. Classification helps us to put together different items according to the feature items decided as interesting and useful. In this paper, we compare two classification methods Naive Bayes and ADTree use to detect spam e-mail. This choice is motivated by the fact that Naive Bayes algorithm is based on probability calculus while ADTree algorithm is based on decision tree. The parameter settings of the above classifiers use the maximization of true positive rate and minimization of false positive rate. The experiment results present classification accuracy and cost analysis in view of optimal classifier choice for Spam Detection. It is point out the number of attributes to obtain a tradeoff between number of them and the classification accuracy.

**Keywords :** classification, data mining, spam filtering, naive bayes, decision tree

**Conference Title :** ICMCSSE 2016 : International Conference on Mathematical, Computational and Statistical Sciences and Engineering

**Conference Location :** Tokyo, Japan

**Conference Dates :** May 26-27, 2016