AN EVALUATION OF NAIVE BAYES CLASIFICATION ALGORITHM IN SPAM FILTERING

Elena Canaj[1], Nelda Kote[2], Odeta Peterci[3]

[1]Department of Computer Science, Polytechnic University of Tirana,Albania,
Email; elenacanaj@gmail.com
[2] Department of Computer Science, Polytechnic University of Tirana,Albania,
Email: neldakote@gmail.com
[3] Department of Computer Science, Polytechnic University of Tirana,Albania,
Email:odetapeterci@hotmail.it

Abstract

This paper treates spam filtering as a text classification process in the context of naive Bayes algorithm. The aim of this paper is to analyse and evaluate through experiments the implementations of three different mathematical forms of naive Bayes algorithm. The three Naive Bayes versions that are evaluated are: Multivariate Bernoulli Naive Bayes, Multinomial Naive Bayes with TF atributes and Multinomial Naive Bayes with Boolean atributes. Many experiments are performed in WEKA environment to evaluate and identify the version that improves the classifications of emails. The results obtained show that the naïve.Bayes algorithm is more performantin classification when using Multinomial distribution with Boolean atributes.

Keywords: *Naive Bayes, text classification, spam, ham, filter*