Paper No: SC-26 **Smart Computing**

Personalized Classification of Non-Spam Emails Using Machine Learning Techniques

Harsha Dinendra*
Faculty of Graduate Studies
University of Kelaniya, Sri Lanka
hdinendra@gmail.com

Chathura Rajapakse
Faculty of Graduate Studies
University of Kelaniya, Sri Lanka
chathura@kln.ac.lk

P. P. G. Dinesh Asanka
Faculty of Graduate Studies
University of Kelaniya, Sri Lanka
dasanka@kln.ac

***Abstract* - With the advent of computer networks and communications, emails have become one of the most widely accepted communication means, which is faster, more reliable, cheaper, and accessible from anywhere. Due to the increased useof email communications, day-to-day computer users; particularly corporate users, find it cumbersome to filter the most important and urgent emails out of the large number of emails they receive on a given business day. Enterprise email systems are able to automatically identify spam emails but still,there are many non-urgent and unimportant emails among suchnon-spam emails which cannot be filtered by conventional spamfilter programs. Though it may be feasible to set up some staticrules and categorize some of the e-mails, the practicality and sustainability of such rules are questionable due to the magnitude of such rules, and the validity period as such rules may become redundant after some time. Thus, it is desired to have an email filtering system for non-spam emails to filter unimportant emails, based on the user's past behaviour. Despitethe availability of research on identifying spam e-mails in the area of further classifying the non-spam e-mails, is lacking. Thepurpose of this research is to provide a machine learning-basedsolution to classify non-spam e-mails considering theimportance of such e-mails. As part of the research, several machine learning models have been developed and trained usingnon-spam e-mails, based on the personal mailbox of the first author of this research. The results showed a significant accuracy, particularly with a decision tree, random forests and deep neural network algorithms. This paper presents the modelling details and the results obtained accordingly.**

Keywords - logistic regression, non-spam email classification, Random Forest algorithms, supervised learning, Support Vector Machines