

TEXT MINING OF EMAILS USING MATLAB CODES

VADIRAJA, N

Assistant Professor – Statistics, Mysore Medical College and Research Institute, Mysuru, Karnataka, India

ABSTRACT

As an application to text mining, it is developed that a Statistical SPAM Filter using Matlab codes. SPAM is defined as 'unwanted Emails'. This is very useful when a workstation receives more number of emails including SPAMs into the email box. Question is how to get rid off from SPAMs. More the SPAMs, greater is the chance of ignoring HAMs (Legitimate Emails). SPAM filters classify emails into SPAMs or HAMs. However, misclassifying hams is worse than misclassifying SPAMs. As a result, it is interested to get rid off from SPAMs keeping in mind that no legitimate emails are ignored. Existing, now-a-days, lots of rule based SPAM filters, however our interest is to develop statistical filter that uses Bayesian and SBPH approach on emails and to compare. Matlab codes are used to generate the test programme.

KEYWORDS: Statistical SPAM Filter Using Matlab Codes