



# Advanced E-mail Spam Detection Methodology by the Neural Network Classifier

V. BRINDHA<sup>1</sup>      J. GEORGE CHRISTOBER<sup>2</sup>      V.NANDHINI<sup>3</sup>

<sup>1</sup>Asst. Prof, Dept of IT, Vivekanandha Institute of Engineering and Technology for Women, Anna University, Tamilnadu, India

<sup>2</sup>PG Scholar, Dept of CSE, Sri Eshwar College of Engineering, Anna University, Tamilnadu, India

<sup>3</sup>Asst. Prof, Dept of CSE, Vivekanandha Engineering College for Women, Anna University, Tamilnadu, India

<sup>1</sup>brindha02me@gmail.com, <sup>2</sup>iamgeorgepaul@outlook.com, <sup>3</sup>nandhinigenuine@gmail.com

---

*Abstract-- Spam is an Commonly said unwanted mail through the internet. In this trend Spam is became an major issue of this world to control legitimately. The large amount of spam will affect the bandwidth problems and mere nuisance to be considered. The issue in these spam mail is to mixed with the new mails and categorization is difficult to produce and it is very tough task to categorize the spam mails and legitimate mails. Spam Filtering is the processing of email to categorize the specific criteria. An anti-spam filtering will be introduced for the classification of Spam mails in the inbox and the spam detector trained by the pre classified emails. The detector detects whether the mail is an ordinary mail or spam mail. Here we include the advance technique called as Neural Network Classifier in the filtering technique to calculate the probability of Spam message from the machine learning from the Spam mails. Neural Network is very accurate efficient, adoptable and robustness and Effective anti-spam approach to classify whether the mail is an Spam or an Ordinary Emails.*

*Keywords--- Spam filtering, Neural Network, AntiSpam, Classifier, Preclassified Mails*

---

## I. INTRODUCTION

The email is increasing nowadays because of their popularity and The Many Industries are found out the email is an easy way to spread out their advertisements to the large number of users at the Very less amount of cost. The Unnecessary emails or unwanted junk mails are called as the Spam mails. The spam mail Majorly contains various Advertisements, Commercial promoters services and herbal products, Insurance advertisements, Hotel reservation and software product advertisements as well as spreading rumours through the mails. The spam mails are growing rate is very high compared to past decennary. It has major issue of normal users, business users, and network administrator. A study of September 2007 reported the 20% of the incoming

messages are spam mails through the internet. It can be reduce the activity of users by time resources and large amount of spam mails will reduce the bandwidth of the network. Hence the project mainly focuses on the email classification by the advanced technique called neural network classification in the anti-spam filtering technique. The classifier will be implemented and designed in the spam email detector and the results are analyzed in the system

## **II. SPAM FILTER DESCRIPTION**

In addition, The spam mails consumes the large amount of storage space, bandwidth and slow down the processes and weaken the mail server because of spam mails and it is also harmful for the computers because the spam mail can also distribute the harm full viruses through the mails even phishing as well through the network. Hence the antispam techniques has been introduced and filtering the front end activities as such as easy way to classify the spam mails earlier in the inbox. The antispam technique has various modules and preclassified emails should be trained in the Neural Network classifiers. In this approach should be easily identifying the important mails and it can detect the spam mails and filter it. The neural network Classifier is one of the advanced machine learning techniques and it should produce the accurate results to finding the spam messages compared with any other classification methods.

## **III. MODULES**

### **MODULES**

- A. User interface Design
- B. Mail Client
- C. Mail Server
- D. Neural Network

### **A. USER INTERFACE DESIGN**

Design a Graphical user interface part for user interaction with application for e.g. user need to register their details includes user name, password, and personal details. After successful registration, user need to login with user name and password. User has three type of process with application that is Compose mail, Inbox, Spam.

### **B. MAIL CLIENT**

Clients are PC workstations on users run applications. Clients rely on servers for resources, such as files, devices, and even processing power. According to process user need to enter the details and send to the server as request. The server provides mail access to that user. After successful login, client checks the inbox or spam or compose mail to another client with the username respectively. Client can send the message or attachment to another client.

### **C. MAIL SERVER**

According to client request server stores the records in the database, for every time the client login its checks the client authentication by comparing the records in its database. Server can play a role to transfer the mail between clients for filtering; classifying the mail server can use naive bayes algorithm steps.

#### D. NEURAL NETWORK

The following steps characterize the whole process:

##### 1. Building a list

- Take the sample spam messages.
- Build a list of words (along with their number of appearance) occurring in all 100 messages, sort it.
- Chose a handy set of words from the top  $n$  of the word list, eliminate domain names etc.

##### 2. Creating the neural network

- Use the top  $n$  word list as input neurons
- Add one hidden layer with as many neurons as input neurons
- Create two output neurons
- Connect every input neuron with every hidden neuron, and every hidden neuron with the two output neurons

##### 3. Train the network

- Create a training set: take the spam emails from the wordlist-set and add ~100 normal, personal spam mails.
- Count in each message the occurrences of the words in the word list and set the according input neuron to this value.
- Activate the network and adjust the weights of all connections using an algorithm like Back propagation, in funtion of the difference between desired and actual output neuron value.
- Repeat these steps until all messages in the testing corpus have been used for training

#### IV. SYSTEM FLOW AND NEURAL NETWORK CLASSIFIER PROCESS

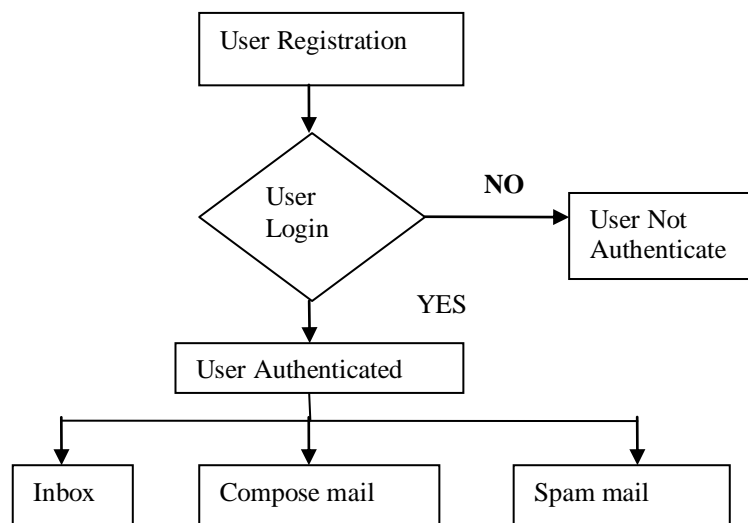


Fig 1. System Flow diagram

System flow diagram represents user need to register details and get the user name and password with that get into login process and verify if authenticated user can able to use compose mail, inbox and spam else not authenticated.

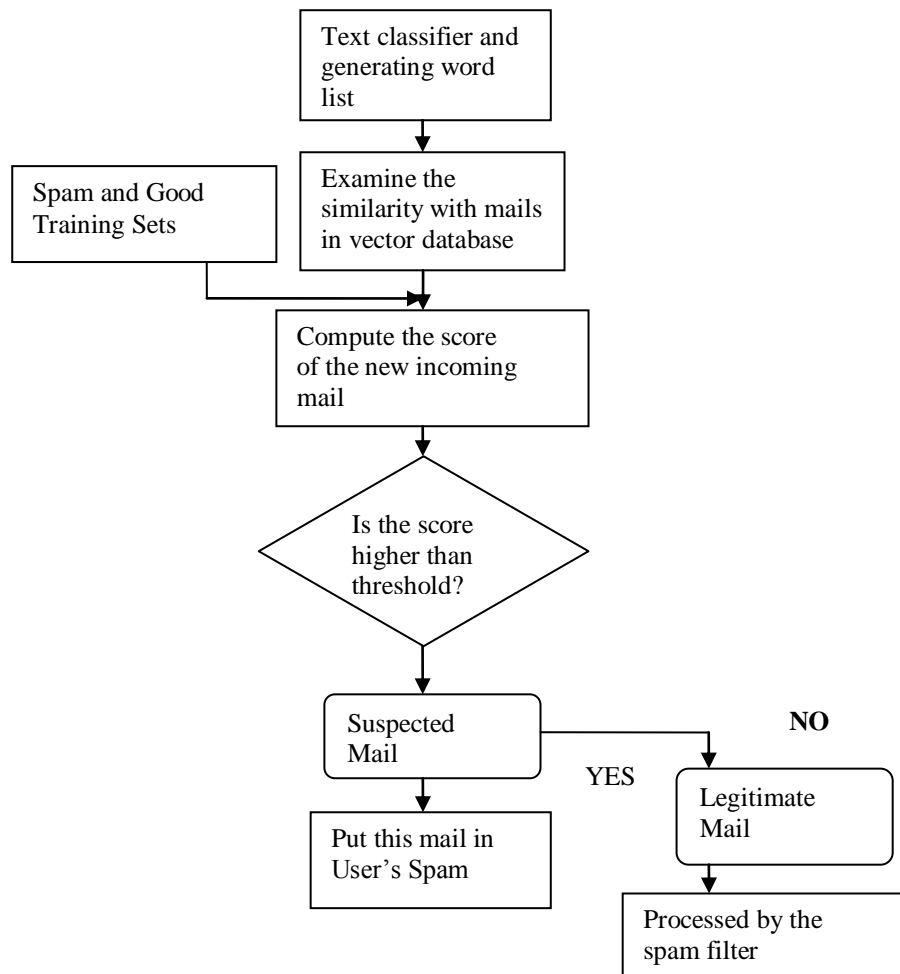


Fig 2.Neural Network Classification Process

## V. IMPLEMENTATION RESULTS



Fig 3. Starting the Server Process

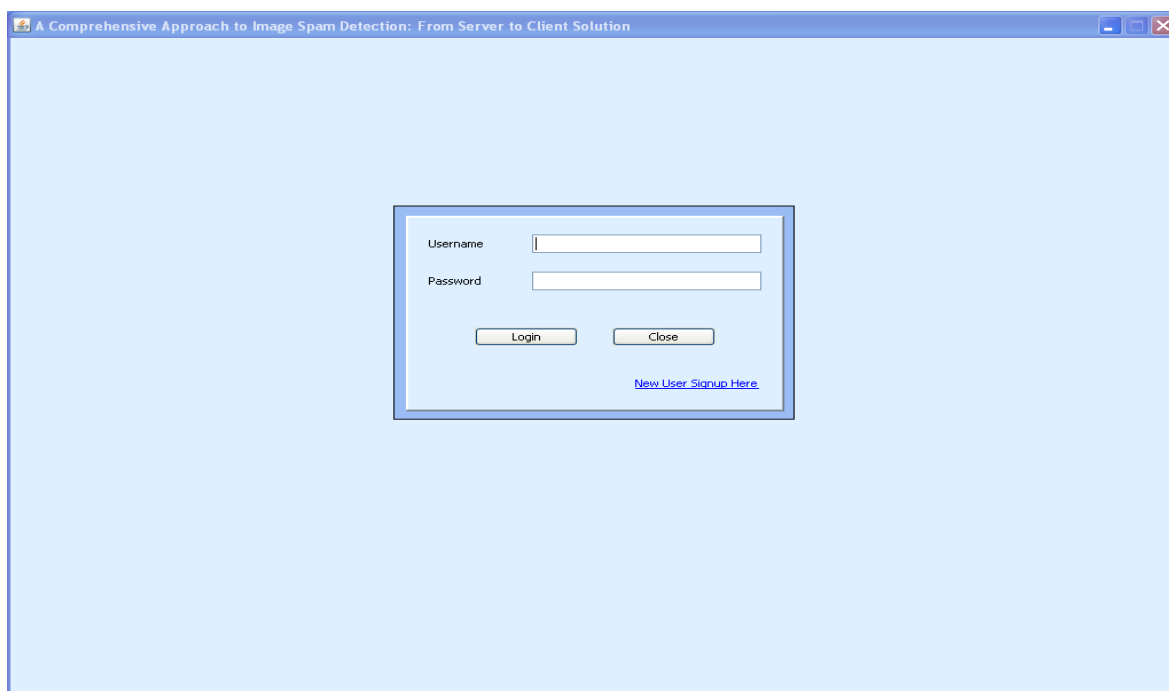


Fig 4. Login and Authentication Process

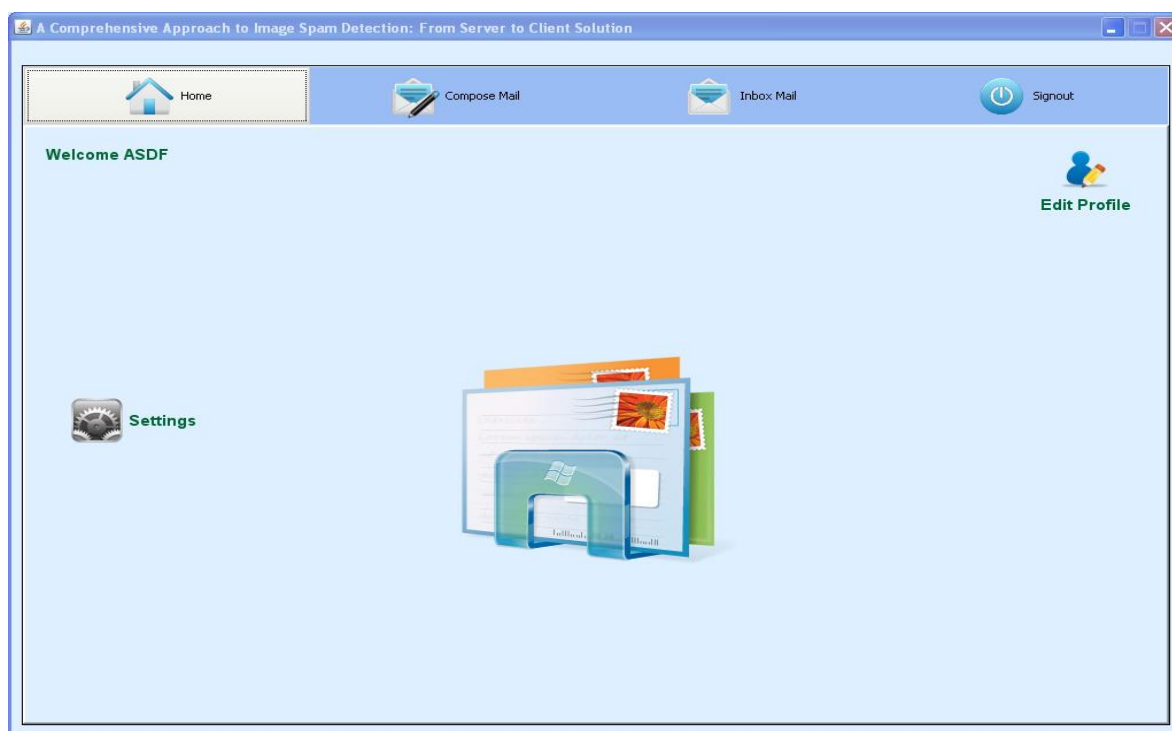


Fig 5 The Main form contains Compose and Inbox Function

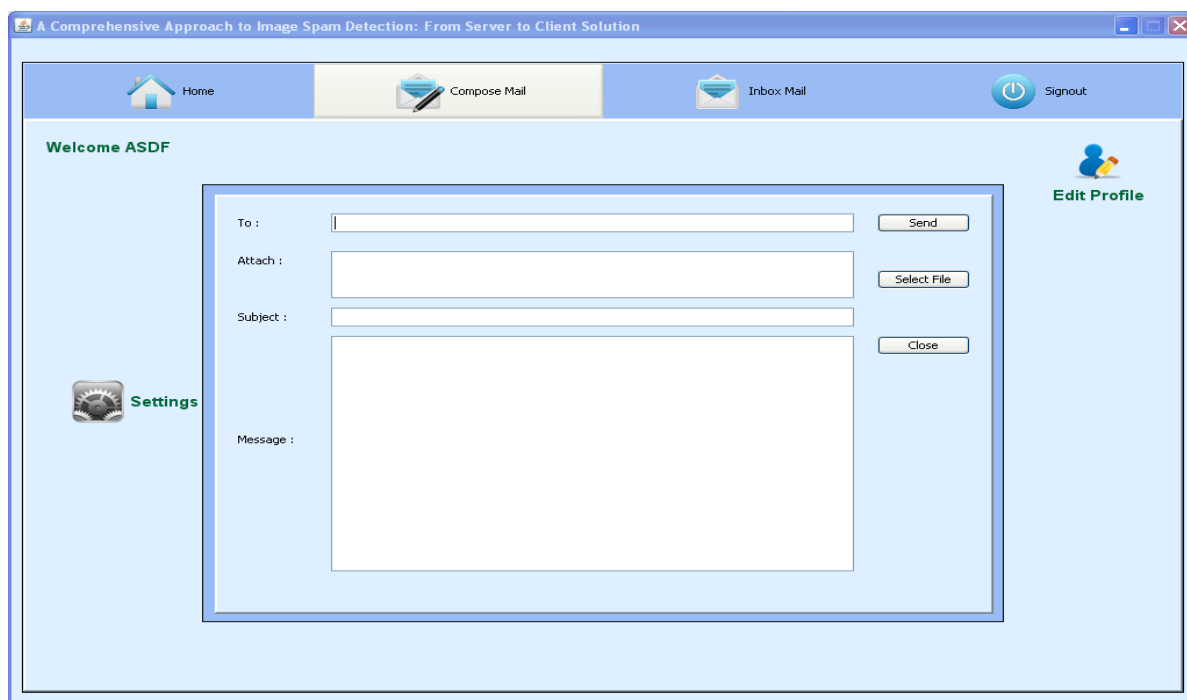


Fig 6 The compose Mail form and it should Authenticate the Spam messages

## VI. CONCLUSION

The main objective of the paper is to apply Neural Network Classifier algorithm on email classification. Also, explored methods used for text pre-processing and probability computation in text classification. Through the project, have learned the advantage and easy to use feature of Neural Network classifier. The scope for improving the performance of the spam email detector. Spam email detectors discuss the potential improvement.

## REFERENCES

- [1] Scott Oaks, Henry Wong, Mike Loukides, '*Java Threads Java Series*', O'Reilly & Associates.
- [2] Patrick Naughton, Herbert Schildt, '*Java™ 2: The Complete Reference*', Third Edition, Tata McGraw-Hill Publishing Company Limited.
- [3] Eugene Charniak, AI Magazine (1990) '*Naive Bayes Networks without Tears*' .
- [4] David Flanagan (1997) '*Java in a Nutshell*', 2nd Edition, May.
- [5] Mark Grand and Jonathan Knudsen (1997) '*Java Fundamental Classes Reference*', Tata McGraw-Hill Publishing Company Limited, 1st Edition, May.
- [6] Drucker, H. and Vapnik, V.N. (1999) '*Support vector machines for spam categorization*', IEEE Transaction on Neural Networks, vol. 10, no. 5, pp. 1048 – 1054, Sept.
- [7] Trudgian, D.C. (2006) '*Spam Classification Using Nearest Neighbour Techniques*', In Proc. of the Fifth Int. Conf. on Intelligent Data Engineering and Automated Computer Science, University of New Brunswick. Full Naive Bayes Network Classifiers, 23 rd International Conference on Machine Learning, Pittsburgh.

## **Authors Bibliography**



**V.Brindha** born in Erode, Tamilnadu, India in 1989. She received B.Tech Degree - Information Technology in Vivekanandha College of Engineering for Women from Anna University, Coimbatore, India. She completed M.E Degree-VLSI Design in Sri Eshwar College of Engineering, under the Anna University, Chennai, Tamilnadu, India. She working as an Assistant Professor in the Department of Information Technology in Vivekanandha Institute of Engineering and Technology for Women, Anna University, Chennai, Tamilnadu, India. Her research interests includes, Network Security, Image Processing and Cloud Computing.



**J. George Christofer** born in Tirunelveli, Tamilnadu, India in 1991. He received B.Tech Degree in Information Technology from Anna University, Coimbatore, India. He is pursuing M.E Degree in Computer science and Engineering in Sri Eshwar College of Engineering , Anna University, Coimbatore, Tamilnadu, India. He is a member of an IEEE Association. His research interests include, Image Processing, Network Security and Web Technology.



**V.Nandhini**, born in Salem, Tamilnadu, India in 1990. She completed her UG in Vivekananda Institute of Engineering and Technology for Women, Elayampalayam and completed her PG in Sona College of Technology, Salem. She working as an Assistant Professor in the department of CSE in Vivekananda Engineering College for Women, Sankari,Salem,Tamilnadu. Her research interests include, Network Security and Web Technology.