ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

Misclassification Error Detection in Email – A Study

P.Yasodha

Assistant Professor, Department of Computer Science, Sri GVG College for Women, Udumalpet, India

ABSTRACT: This research paper primarily explains about the comparison of algorithms and methods used to reduce misclassification of error detection in email. The importance of this paper is to analyse two different algorithms and find the best accuracy in it. The paper generally explains about the overall study of Decision Tree Algorithm and Naïve Bayes Theorm which gives an optimized system performance.

KEYWORDS: Misclassification , Bayes's classifier, Decision tree.

I. INTRODUCTION

In general people communicate a lot through email where highly secured documents and data are shared each and every day .In particular an email is a combination of text, graphical message and sound files. The main issue in email is that with the security and large amount of spam in inbox , which the user needs to delete in particular. Email is an encoded ASCII text through each text is drawn using text editor which are inturn broadcasted using text and multimedia files,which is called email server which causes an increasing possibility of viruses and spams.

II. RELATED WORK

The reduction of misclassification is done by formulating legitimate emails which categorise into junk mails. This paper helps to differentiate cost among algorithms for classifiying junk mai ls over legitimate mails and vice versa. The basic cost classification is done by two different algorithms

1. Naïve Baye's classifier
2. Decision tree algorithm

In general linear discrimaint analysis is initated to process the cost of error detection . Basically two different data were used for classification , training data and sample data In training data , an observation is done known class label from which sample data are used . A misclassification error is computed by resubstitution error on training data. The term confusion matrix is contained about class labels and predicted class labels is computed in training sets.On the whole Quadratic Discriminanat Analysis is preferable.

Naïve Baye's Classifier

It is an classification technique based on bayes theorm , which shows the individuality of data sets.In Naïve Baye classifier, any unrelated large data sets can be classified [1]. The advantage would be multi class prediction with less data node which perform well in numerical categorical values.On the other hand, this classifier is a bad estimator[9]

Decision Tree Algortihm

The algorithm is based on decision support system which is operated using graph decisions with tree –like representation , which results in reduced cost and utilities. A decision tree is a flow-chart like structure where test is conducted on each node[2]. This algorithm is based on a classification tree in which the operation on deducing values of dependent attributes to independent attributes . A classification rule is classified in paths initiated from root to roof. A decision tree algorithm is divided into two types as categorical and continious variable These kind of algorithms are used in operational research , decision analysis, easier classification [2].

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

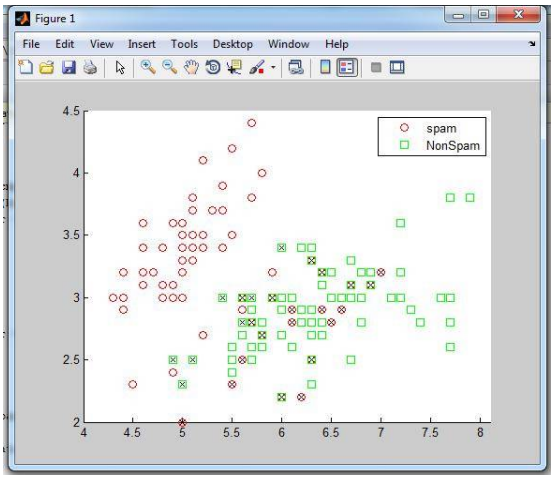
(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

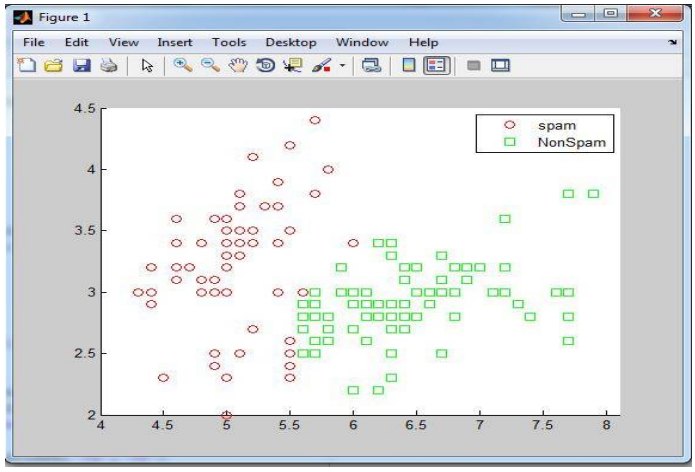
III. EXPERIMENTAL RESULTS

The Naïve Baye's classifier works on the principle of datasets with limited resources in terms of CPU and Memory. Each classification on naïve is independent , but our paper need a comparative analysis on large data sets with unique analysis over different classification. This would be probably gained by Decision tree algorithm where , an easier study is done literally on infinite datasets with minimum cost . In decision Tree algorithm , an data exploration is done simply and effective manner.

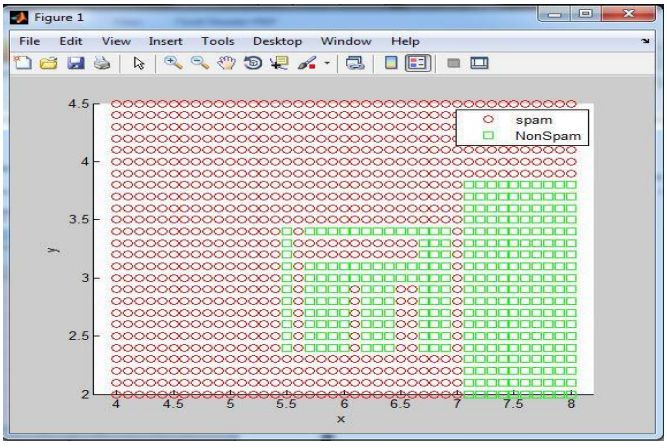
The figures shows the comparative result for using datasets on the labels of spam and Non Spam by using different classification .Figure1 shows the Scattering of the dataset on the basis of the class labels spam and Nonspam . Figure 2 shows the Misclassification of datasets in spam and Non Spam . Figure 3 shows the Classification using Naive Bayes Gaussian distribution. Figure 4 shows the Scattering of decision tree based evaluation. Figure 5 shows the classification plotted using decision tree classifier. Figure 6 shows the visualization of the classified dataset using WEKA tool.



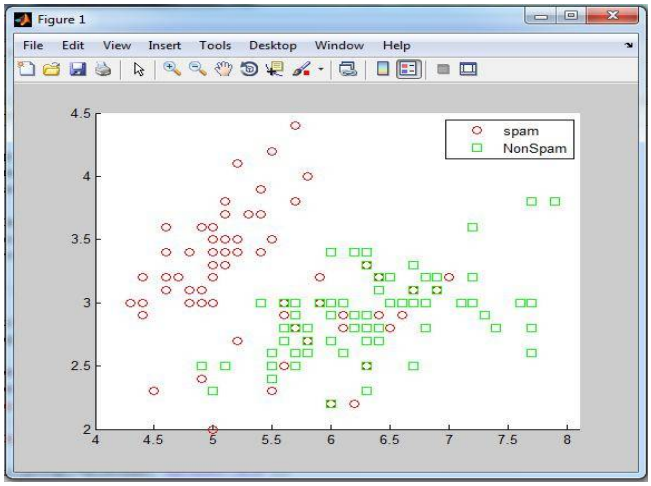
(1)



(2)



(3)



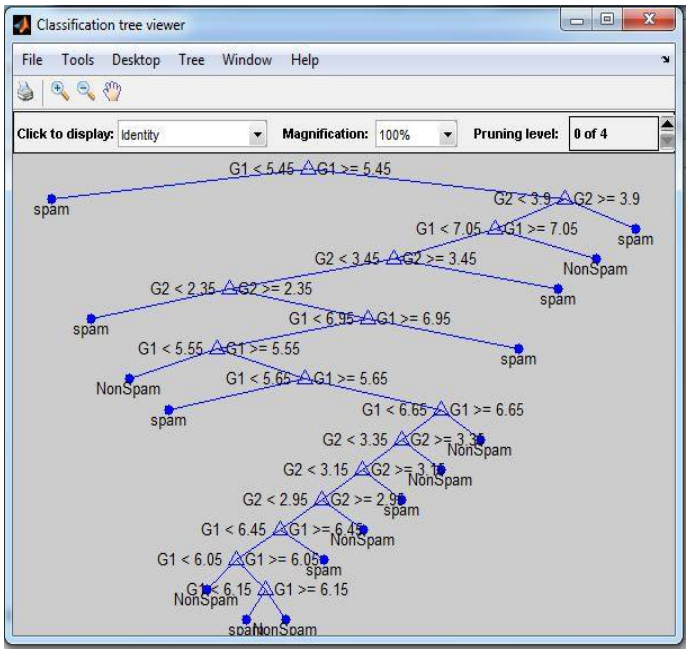
(4)

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

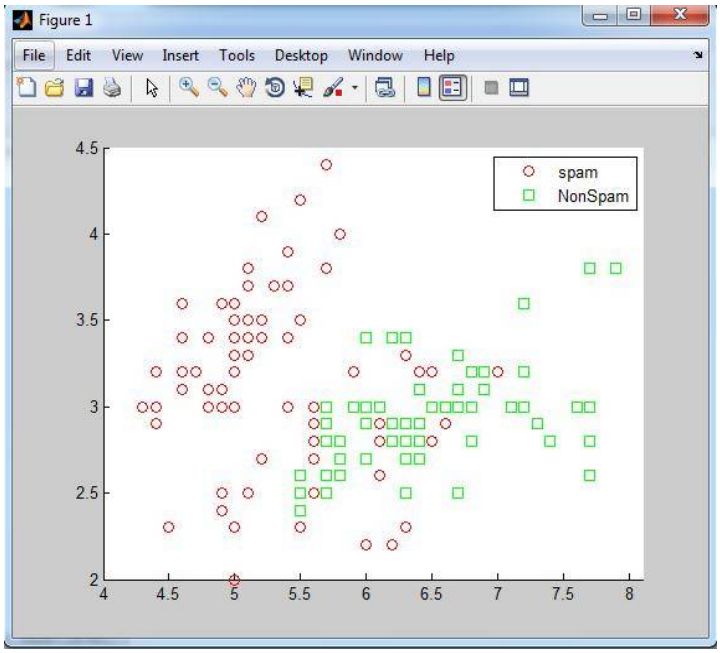
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

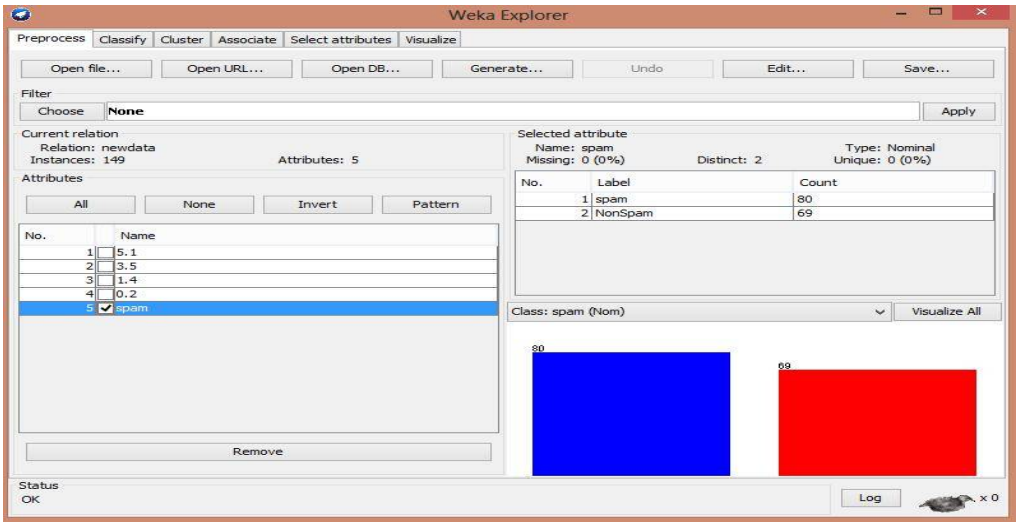
Vol. 5, Issue 9, September 2016



(5)



(6)



(7)

VI.CONCLUSION AND FUTURE SCOPE

Filtered mails are being filtered to measure the misclassification using different data mining technique . By using proper analysis and experiments , decision tree algorithm is the best classifier and easier to interpret with different number of datasets. In this paper , two different tools are used , WEKA tool and mAT Lab. In future , the attributes of data sets can be further classified in advanced network technology.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

REFERENCES

[1] Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Sakkis, G., Spyropoulos, C. D., & Stamatopoulos, P. (2000). Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach. arXiv preprint cs/0009009.
[2]Jigna Ashish Patel ,"Classification Algorithms and comparision in Datamining "- International Journal of Innovations & Advancement in Computer Science May 2015.
[3] Rokach, Lior; Maimon, O. (2008). Data mining with decision trees: theory and applications. World Scientific Pub Co Inc.
[4] Tan P-N, Steinbach M, Kumar V (2006) Introduction to data mining. Pearson Addison-Wesley
[5] Rokach, L.; Maimon, O. (2005). "Top-down induction of decision trees classifiers-a survey". IEEE Transactions on Systems, Man, and Cybernetics, Part C 35 (4): 476–487.
[6]Classification algorithm in Data mining: An Overview IJPPT-2013 Vol.4 issue 8.
[7]http://blog.datumbox.com/machine-learning-tutorial-the-naive-bayes-text-classifier
[8]https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained.