Thesis: Title: Use of artificial intelligence and data analytics to compare the e-mail spam filtering (YAHOO, Gmail, Outlook) depending upon user experiences

Chapter # 1: Introduction

Student Name: Shoaib

Table of Contents

[Introduction 3](#_Toc88226008)

[Machine Learning & Deep Learning Method 4](#_Toc88226009)

[Novel Email Spam Detection Method 5](#_Toc88226010)

[Spam Email Detection Using Machine Learning & Neural Networks 6](#_Toc88226011)

[Machine learning algorithm on email spam classification 7](#_Toc88226012)

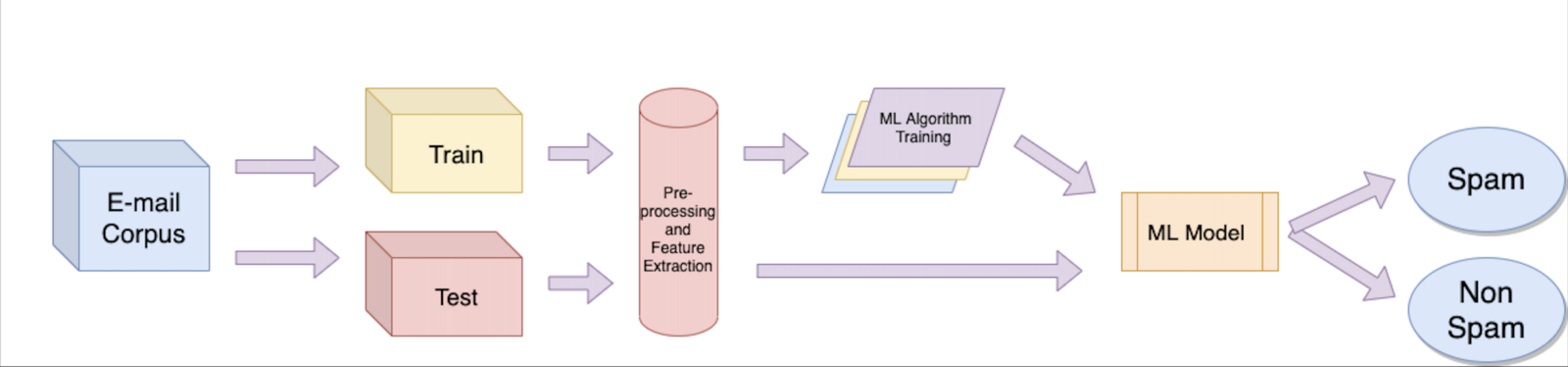
[Aim & Objective of this Research 8](#_Toc88226013)

[References 9](#_Toc88226014)

# Chapter # 1: Introduction

# 

Email spam is major issues for the enterprise organization because spam mail contains malware information that executes on the operating system of the use and causes problem (Alzahrani & Rawat, 2019, ) spammers and hackers sends various spam emails to the user and steal the essential user data and information. This problem causes an issues to resolve the spam and hack email at the enterprise level the machine learning method used to meeting this challenge. Famous internet service providers including email service provider such as Yahoo, Gmail Outlook uses machine learning method to filter and clean the spam email. It only put the spam email separate email folder, it does not able to detect and delete the spam mails automatically. So the enterprise organization using email server to meet these challenges, being the support of network administrator, the deploy machine learning algorithm in the email server because email server does not facility of machine learning method to detect filter and clean spam email and spam messages in the organization. The development of machine learning AI algorithm in email server that helps and support the corporate to eradicate the spammers attack and hacker attack. The short message is most popular in these days because short sms service gain popularity to build up the policy of mobile communication architecture. Various advertising agencies advertised products using the short messages method, send the unwanted information continuously to the user it might be busy internal memory of system and might be able to steal the data and hack the operating system. The spam messages produces irritation, problem and frustration them it also engaged the user for deletion and reading activity, it causes the problem that wasting the user time and user memory of internal operating system, it might be also able to cost the memory.



*Figure 1: E-mail spam and non-spam filtering*

The popular machine learning algorithm method successfully deployed in this thesis to come up the challenge of spam email and spam messages. The essential machine learning algorithm used to detect filter and clean the spam email data such as logistic regression, naïve bays, support vector machine, and neural networks, the method filter the messages and measure the accuracy of most effective machine learning algorithm that deal on spam messaging technique, beside this neural network performs better accuracy due to hidden neural network layer, because neural networks performs well to trained the classifier model and used the method to detect the harm or spam email.

## Machine Learning & Deep Learning Method

(Tazmina, et al., 2020) Deep learning works on efficient neural network layer, neural network contains billions of hidden neural layer. The deep learning algorithm works very effective to detect the spam images in the email dataset. The using of convolutional neural network that achieve the highest accuracy by using the image dataset. Beside this machine learning method also works very well by using logistics regression decision tree, random forest tree, naïve bays that achieve the highest accuracy of spam email detection including 98% to 99% of accuracy with F1 score. The uninvited bulk emails which contains malware information and malware dummy link that causes the problematic situation to analyze the text based situation. The image segmentation method widely used to detect the unwanted emails. The text based analysis to detect spam text in large dataset file is also highlighted and deep learning method able to detect the spam text over the large dataset file.



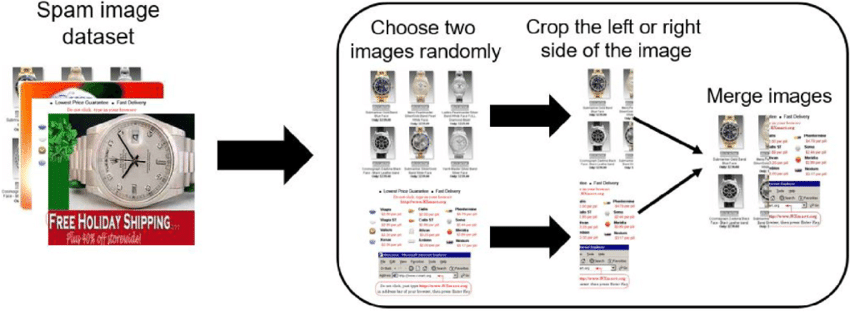
*Figure 2: Deep Neural Network*

## Novel Email Spam Detection Method

(Ezpeleta & Mendizabal, 2020) Unwanted junks spam email is part of every email domain that very difficult to tolerate them at enterprise level. In previous research various machine learning method has been used to detect and filter the spam emails. The analysis method to detect the efficiency of detection the spam mails the hypothesis framework deployed to test the working efficiency of machine learning algorithm. The sentiment analysis which is based on detect the personality recognition which extensively used the email account, various hacker and spammers target that audience which does not know about the spam email and spam short messages. (W. Peng, 2018,) Enhancing the naïve bays spam filter through intelligent text modification detection, since the issue exist to detect and clean the spam email at the enterprise level, various industries application using the corporate email server to store and retrieve email data on particular server computing, so the email server does not able to detect the spammer. The email correspondence has been increase so far due to all official work required email communication which needs some tools and technique to do the various official task and need approval, so at this level urgent need of spam filter that detect, spam and eradicate it automatically.

Spam email stores in user inbox, and also change the text of original mails due to malware and phishing attack served in email system. Currently the naïve bays machine learning algorithm effectively works very well to detect and refine the spam email from the user inbox. The implementation of novel algorithm that works with naïve bays which produces highest accuracy result that correctly detect the spam and ham. This project is going to use python programming framework which is most popular programming platform of machine learning and deep learning algorithm development discovery.

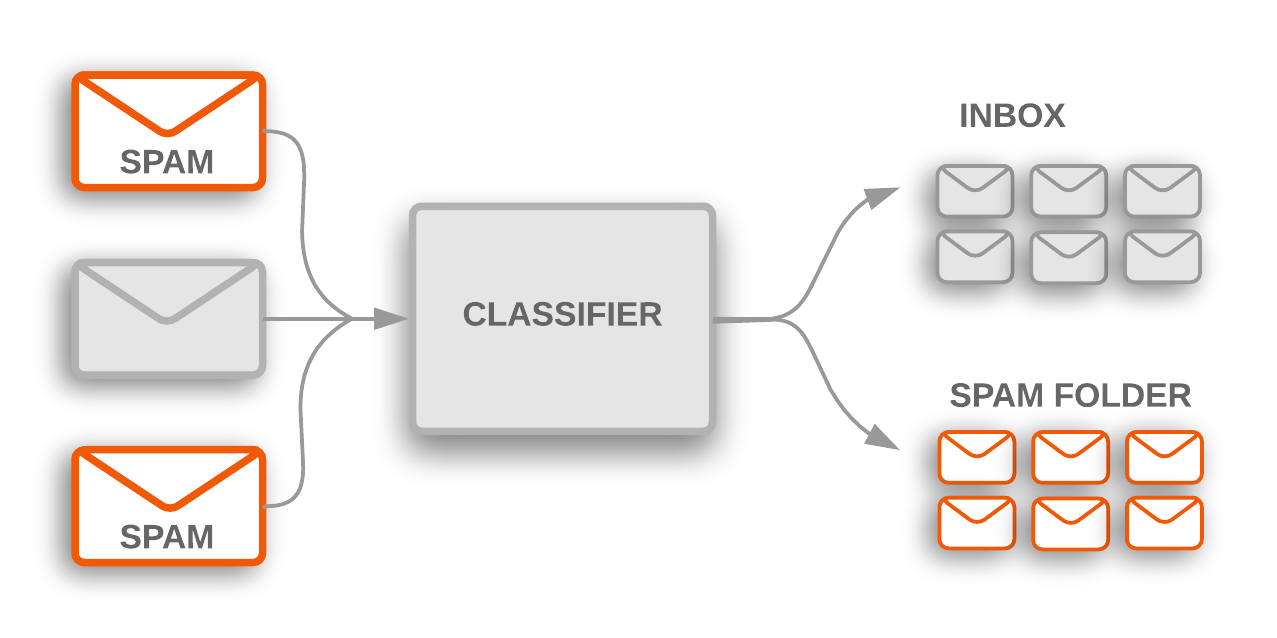
Python extensive works very well to detect various machine learning discovery, data science, big data analytics and artificial intelligence taking initiative to develop various AI based python libraries which essential works very well comparative to other programing language.



*Figure 3: data augmentation process for spam image detection*

## Spam Email Detection Using Machine Learning & Neural Networks

(Sethi Sumesha, et al., 2021) Spam emails known as junk email which does not need approval to enter in the user inbox, the spam messages contains with spam hacker id that enter in the corporate to get the secret information and important asset details. The area of spam detection which is considered to be very limited due to some limited types of domains and networks. Gmail is considered as one of the best email service in the world, it also helps to provide domain & network facility to enterprise business. Google built in with enhance cyber security features that support various domain and cloud computing also. Since the Google networks inspires from Yahoo, but the server and searching index of Google contains various types of security level it also support the Gmail to protect the end to end encryption for message sending and message receiving to the user. Particularly machine learning method extensively used to detect the spam and ham emails, there are mainly two features has been introduce such as stopwords and wordcount these method analyze the email either the email is spam or ham.



*Figure 4: Spam Classifier*

The entire email classification process is called spam detection which used to protect the user personal information and organization data. Various machine learning algorithm works efficiently to feature the spam and ham emails, various spam email contains malware attack link that used to destroy the organization server and networks computing and it also destroy the confidential data. Machine learning algorithm multinomial naïve bays, logistics regression, linear support vector machine including artificial neural network algorithm that used to determine the spam emails. Dataset has been used in this project from kaggle and Github repository. Detection of spam and ham emails and spam text and it detect the malware emails, malware emails used in previous windows releases, but the window 10 operating system updated.

## Machine learning algorithm on email spam classification

(Neha., 2020) The message sending towards the user computer to other is considered sensitive matter due to some security breaches. Email classification method used with machine learning algorithm that works programming platform to manage the spam and ham messages over the internet usages. The research is focused on email spam detection beside this mobile communication short messages is also discussed and corporate large file text included with spam text has been highlighted. Unwelcome email coming from hackers site, it may use to spoil the communication network and destroy them. The development of deep learning approach with neural network that effectively works to detect spam image by using the image dataset of spam image and clean image dataset. Various dataset has been provided on Kaggle and Github repository which support various research to deploy spam filter method to avoid any issues and problem in the organization. The objective of this research to conduct compressive machine learning algorithm technique such as SVM, DT, KNN, NB, LG, RF, LR CNN which combines all of these method in this research that support the research to deploy the email spam detector algorithm required in premises.

The rest of the thesis planned as follows in chapter 2, the comprehensive literature review and related work has been conducted, which is known as secondary research data, in chapter 3 the methodology framework of this research is planned and discussed including research paradigm and research philosophy, and the methodology of machine learning algorithm that perform the spam detector which provides in chapter 4 the analysis of machine learning algorithm such as SVM, DT, KNN, NB, LG, RF, LR CNN also evaluated with their efficiency and level and various spam filtering technique is also evaluated and critically analyzed with other researcher. Chapter 5 offers the discussion in which the frameworks of the overall structured machine learning algorithm benefits, advantages, drawback discussed, future requirement and challenges is highlighted. And at the end conclusion is described.

# Aim & Objective of this Research

* The objective of this research to combine all machine learning algorithm to detect spam email and spam text and spam messages in the dataset.
* The aim of this research to develop python programming applications which executes on server platform and helps the organization to detect and filter the spam emails and spam text.
* The goal of this research used the secondary source of data from Kaggle &Github repository, and the goal of this research to protect the confidential email of the user.
* The main goal of this research to block unwanted emails, block unwanted spam links and block malware phishing attack emails.

# References

Alzahrani, A. & Rawat, D. B., 2019, . Comparative Study of Machine Learning Algorithms for SMS Spam Detection. *2019 SoutheastCon,,* pp. 1-6.

Ezpeleta, E. & Mendizabal, I. V. d., 2020. Novel email spam detection method using sentiment analysis and personality recognition. *Logic Journal of the IGPL ,* 28(1), p. 83–94.

Neha., S., 2020. *A Study of Machine Learning Algorithms on Email Spam Classification,* Missouri State University: Southeast Missouri State University. ProQuest Dissertations Publishing, 2020. 27993789..

Sethi Sumesha, M., Vinayak, C. & Dahiya, C., 2021. Spam Email Detection Using Machine Learning and Neural Networks. *Sentimental Analysis and Deep Learning,* Volume 1408, pp. 275-290.

Tazmina, Fobia & Kateriana, 2020. Convolutional neural networks for image spam detection. *Information Security Journal: A Global Perspective ,* Volume 3, pp. 1-10.

W. Peng, L. H. J. J. a. E. I., 2018,. Enhancing the Naive Bayes Spam Filter Through Intelligent Text Modification Detection. *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), ,* pp. 849-854.