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

Chapter # 5: Discussion & Conclusion

Student Name: Shoaib

Table of Contents

[Chapter # 5 Discussion & Conclusion 3](#_Toc95143595)

[Discoveries 3](#_Toc95143596)

[Novelty 3](#_Toc95143597)

[Conclusion 5](#_Toc95143598)

[References 6](#_Toc95143599)

# Chapter # 5 Discussion & Conclusion

(Zamir & Hikmat Ullah Khan, 2020) Email spam detection method using the diverse machine learning approach to detect spam and ham emails using the email dataset and spam email dataset. Email widely used at enterprise and corporate business level that support the organization multi chain domains. So far the business center facing lot of problems and hurdles to clean the spams and malware emails. Malware email attack is minimize so the spam junction is not eradicate yet. Email server support enterprise email login and email activity facilities it also used to manage the email storage access and allocate the user domain panels.

## Discoveries

Examination and investigational results used various machine learning method such as naïve bays classifier used multinomial algorithm to classify spam and ham emails using the dataset, it achieve the 98% accuracy. Beside this support vector machine and logistic regression effectively deploy with TFIDF algorithm to transform the text into numeric array and evaluate the each array cells in terms to investigate the algorithm. Streamlit is also deploy and tested in web based graphical application to detect spam and ham text, background the naïve bays algorithm has been developed that classify and integrate the text into the numeric form and predict the text according to the label based approach. Using these proposed model deep neural networks with sentiment features performed the machine learning classifier in terms of classification accuracy about 97% to 98% of detection.

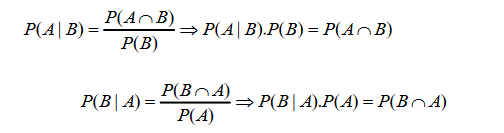
### Novelty

The research is novel research because streamlit concept has been introduced with the help of naïve bays algorithm. Multinomial feature extraction and feature detection algorithm deployed with the help of sklearn library using the python anaconda jupyter notebook.

(Bibi, et al., 2020) With evaluation of various machine learning algorithm on spam email detections outcomes present that the naïve bays algorithm present effective accuracy and precision using the python programming also SVM algorithm present the great score to detect the spam emails by using the dataset.

The term frequency inverse document used effectively to transform the string text into the numeric text conversion to evaluate the spam and ham text by 1 and 0 in terms of effective statistical analysis. TFIDF algorithm also used to count the word and also complement the words in the document its weights the count features.

Naïve bays classifier predict on probabilistic based which is good for text evaluation. Naïve bays called the occurrence of one feature independent of occurrences of other feature, supervised machine learning algorithm to classify the high dimensional dataset. It works on bays theorem.



Naïve bays calculate the conditional probability beside this on the principle of random variable.

Email Classify to Detect Spams the following steps:

1. Training
2. Extract email text
3. Parse each email token
4. Training and testing the dataset with sklearn feature extraction library
5. Filtering the dataset
6. Classify the text based on label

The performance metrics of accuracy evaluated the computed number of digits that process and present the text.



(M, et al., 2020,) Logistics regression works with NLP toolkit natural language processing that process the text with countvectorizer to evaluate and integrate the dataset. Stop word and word count clean the text and clean the white spaces in the text, to measure the text performance it uses the TF-IDF algorithm that convert the text and evaluate them.

(Sharmaa & Yadav, 2021) In these days the spam mails objective has been changed now it’s part of changing the scenario’s , spam is used to hack the confidentially data & information, it also uses as play card to capture the corporate secrete asset information and destroy the computer networks. The objective has been changed and cyber security trying to resolve this kind of issues with development secure socket layer to hide the confidential data. Gmail services support end level security to the user, also encrypt the user data by sending to Google drives storage location. The largest file easily sent to Google drive with private and public access. Gmail services is much better than comparatively analyze to Yahoo, Outlook server. Beside this decision tree, k-nearest neighbors achieve 90% accuracy to detect spam and ham emails, social media fake profile detection decision tree algorithms helps a lot to sort out the fake emails profile id.

## Conclusion

The spam and ham email feature extraction in email dataset, python machine learning algorithms effectively works to meet the challenges and provides the interface to detect the spam emails spam id and spam text in the dataset. Dataset has been downloaded from Github and kaggle repository. Naïve bays classifier algorithm detection frequency is much better than as compare to support vector machine and logistics regression algorithm. Beside this python streamlit web based library to creative & build the spam application and shares with colleagues easily. Support vector machine classifier uses the NLTP, natural language processing mechanism to process the text in terms numeric. Logistics regression works with TFIDF classifier that transform the text into the numeric array and evaluate each bit in term of true and false. This project is going to help other researcher to clearly understand the objectives of email spam classifier and uses this project at corporate level to handle organization emails. Future research required to improve the yahoo and outlook emails server, comparatively Gmail server uses machine learning algorithm and works effectively and better.

### References

Bibi, A., Rasia & Samina, 2020. Spam Mail Scanning Using Machine Learning Algorithm. *Journal of Computer,* 1(1), p. 1.

M, T. M., K. Y. & Thida, A., 2020,. A Comparative Approach to Naïve Bayes Classifier and Support Vector Machine for Email Spam Classification,. *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), ,* pp. 324-326, .

Sharmaa, V. D. & Yadav, S. K., 2021. An effective approach to protect social media account from spam mail – A machine learning approach. *An effective approach to protect social media account from spam mail – A machine learning approach,* Volume 1, p. 1.

Zamir, A. & Hikmat Ullah Khan, 2020. A feature-centric spam email detection model using diverse supervised machine learning algorithms. *Emarld insights,* 38(3), p. 1.