COMP 455

Term Project – Email Spam Filter

Name: Tushar Mahendra

Student ID: 300156940

Submitted To: Gabriel Murray

To: Gabriel Murray

Computer Information System Department

Abbotsford Campus

Subject: Report on Term Project - A Spam Email Filter.

Introduction:

In the term project, I worked on how to filter spam emails which is a big problem of today's world. There are thousands of emails in your mailbox and you've lost important emails in them. Well, here is an algorithm that works closely to sort the spam emails and important emails.

Background:

The Spam Email Filter is important for a user as it prevents unwanted and virus infected emails. It is essential for a user to stop as much spam as possible to protect your network from risk of malicious software, virus attacks and phishing webpages.

This is a content-based spam filter. It scans the words that are commonly used in spam emails. This email spam detection is created using Python programming language which checks the subjects of emails.

Problems:

There were multiple problems faced while working on this project:

- **Less accuracy:** Firstly, I worked with CountVectorizer which was less accurate. Later, I found online that TfidfVectorizer is highly accurate than CountVectorizer. There is still a higher chance that some spam emails might pass through the filter.
- **Finding sample file:** It took a while to find the correct sample file. I used Kaggle.com to find emails.csv which had two columns 'text' and 'spam'. **Text** has

subject and **spam** has ‘1’ or ‘0’. 1’s are spam emails and 0’s are important emails.

The first file I found had email addresses only which wasn’t useful in this program.

Algorithms used:

- Naïve Bayes Algorithm
- TfidfVectorizer

Summary and Conclusion:

This Email Spam Filter protects the server from overloading. It avoids emails that can damage the network, server or computer.

```
In [215]: # We get the accuracy by dividing count by length of prediction.
          count/len(pred)

Out[215]: 0.9113257243195786
```

I hope you’re convinced with my email spam filter. It gives a high accuracy score to avoid unwanted emails.

Reference:

Email Spam Detection Using Python & Machine Learning. Retrieved from:

https://www.youtube.com/watch?v=cNLPt02RwF0

Email Spam Detection Using Python & Machine Learning (August 7,2019). Retrieved from:

https://randerson112358.medium.com/email-spam-detection-using-python-machine-learning-abe38c889855