

**Synopsis for Minor Project**

**Email Spam Detector**

**Computer Science & Engineering**

**(CSE-1/5th Semester)**

**Team Members: Guide Name:**

**Ashuwani Kumar Singh Er. Dimple Bhasin**

Chapter 1

**Introduction** Due to the wide popularity of the internet and its communication with no cost, it was recognized as the premium tool for advertising and marketing. With respect to economic constraints, most number of people started sending emails to thousands of people across the world. This made internet, a commercial network with the association of electronic mail as one of the quick resources of communication. The major problem in today’s internet world is sending bulk or unsolicited emails to numerous users. This adds an additional advantage of launching other attacks and wasting of resources [1]. E-mail spam comes under the electronic spam which sends bulk of unnecessary or junk mail of duplicate emails to the recipients. 1.1 Types of spam: 1.1.1 Email Spam: Email spam is the most familiar spam that most of the users come across every day. Email spam follows three properties i.e., anonymity, mass mailing and unsolicited emails. Anonymity is the property of hiding the uniqueness and whereabouts of the email sender. Mass mailing is defined as the sending of bulk identical emails to the large number of groups and unsolicited emails are the emails transferring to the recipients who do not request. Typically, an email sent to large number of groups without any request by hiding their identity is referred as email spam. [1] [16] 2 1.1.2 Comment Spam: This is the most common spam that many users come across in various blogs. Spammers use the posts in the blog to redirect to spam websites. The ranking of such blogs gets increased gradually in the search engines. It is basically used to promote the searching services like Wikipedia, blogs, guest books etc. There are number of tools in the market to get rid of comment spam. 1.1.3 Instant messenger spam: It is not as widely spread as other types of spam. Yahoo messenger, My space, Windows live messenger etc. are the end spots for the spammers. The spammers gather the data of different users and send unsolicited messages within a link that triggers viruses, spam etc. The best way to get rid of this type of spam is to ignore the messages from the strangers. There is also a possibility to get the links from the existing friends list. A critical measure like verifying the size of the URL will be able to trim the chances of being the victim to this spam. 1.1.4 Junk fax: Junk faxes are not as prevalent as before. It reduced periodically with the existence of internet technology. However, there are also some risk factors occurring in few corners because of this telemarketing technology. This is similar to junk email where the advertisements and messages are passed to numerous users via fax machines. The adversaries use broadcast fax as a medium to pass on the junk fax to various users. Fortunately, there are surplus tools to overcome junk fax. 3 1.1.5 Unsolicited text messages: This is kind of similar to instant messenger spam but here the messages are passed via mobiles. SMS is the service through which the messages are transferred from one user to other user. The easiest way is to maintain the contact with the known friends instead of strangers. It is relatively easy to find the source where the message is coming from with the instant messenger spam. It is critically important no to click on the links that are passed via mobile by the spammers. 1.1.6 Social networking spam: Social networking sites play an important role in today’s world. With the advent of such sites, spammers also started flooding using new techniques to make the social networking sites such as face book, twitter, linked in etc. as part of the spamming activities. As of now it is targeting only the wall posts, messages but these techniques evolve certainly over a period of time. Spammers use notes or messages through various groups or pass the messages with embedded links, which may lead to pornographic or other sites and target spam [4]. Even though these sites have an option to report spam or abuse activities, the spammers frequently change their address or account to hide their identities.

\*Page to be Turned

Chapter 2

**Objectives :**

**Characterizing botnets from Email Spam Records**: This framework presents techniques, which use traces of spam email to map botnet in groups. This is done by viewing for several bots involving in the same spam campaign. This has been used against a sample of spam email from Hotmail web mail system and has successfully detected multiple botnets. This technique uses a large set of spam emails as input, which are destined at Hotmail in a regular period. Group of botnets that involve in launching spam emails with respect to its statistics like dimensions and events are observed in the output. Three major steps are involved in identifying them. 14 With “clustering email messages,” spam email messages with identical content are transferred from the entity which is controllable [12]. Firstly, the spam campaigns are to be identified. Most of the spam content in the campaigns appears to be similar, but in order to evade the differences; identical properties are to be considered. For efficiency, shingling concept is used [2]. A unique property like finger print is considered here and counts the number for every single message. Based on the number of finger prints that share among the spam campaigns identical nature is computed. The second step involves “Assessment of IP dynamics” where the results are dependent on the dynamic IP addresses of the hosts. But, for this approach where the results should affect the dynamic IP addresses, different segment of IP address space are to be used for the computation of IP based parameters. For every C-subnet, two parameters are to be used through which it calculates whether two spam messages occur from the same machine at distinct times. The two parameters include average IP address reassignment time and IP reassignment range. The final step is “Spam campaigns to botnets”, where it is assumed that group of spam email messages is possible to join from the identical spam network. Depending on the first two steps, it combines every spam campaigns into a set of spam campaign that is originated from the same network of bots. Based on the statistics of the IP dynamics, for each corresponding message in a spam campaign, even the sending host involves in the consecutive spam campaigns. Two spam campaigns are also combined if huge numbers of participants are involved. [9]

Chapter 3

**Hardware & software requirements :**

A Computer System/Laptop

GPU

Ram: 4GB

Octa Core Processor

Chapter 4

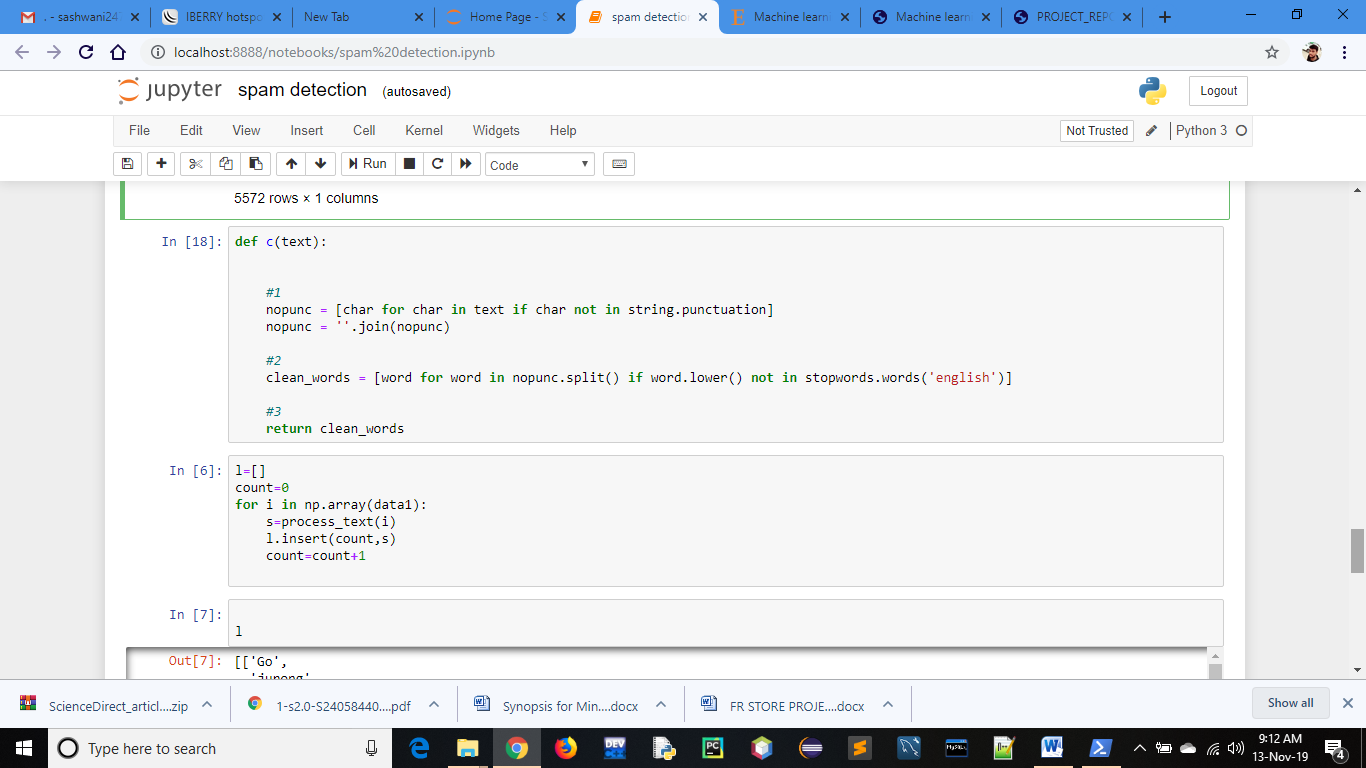
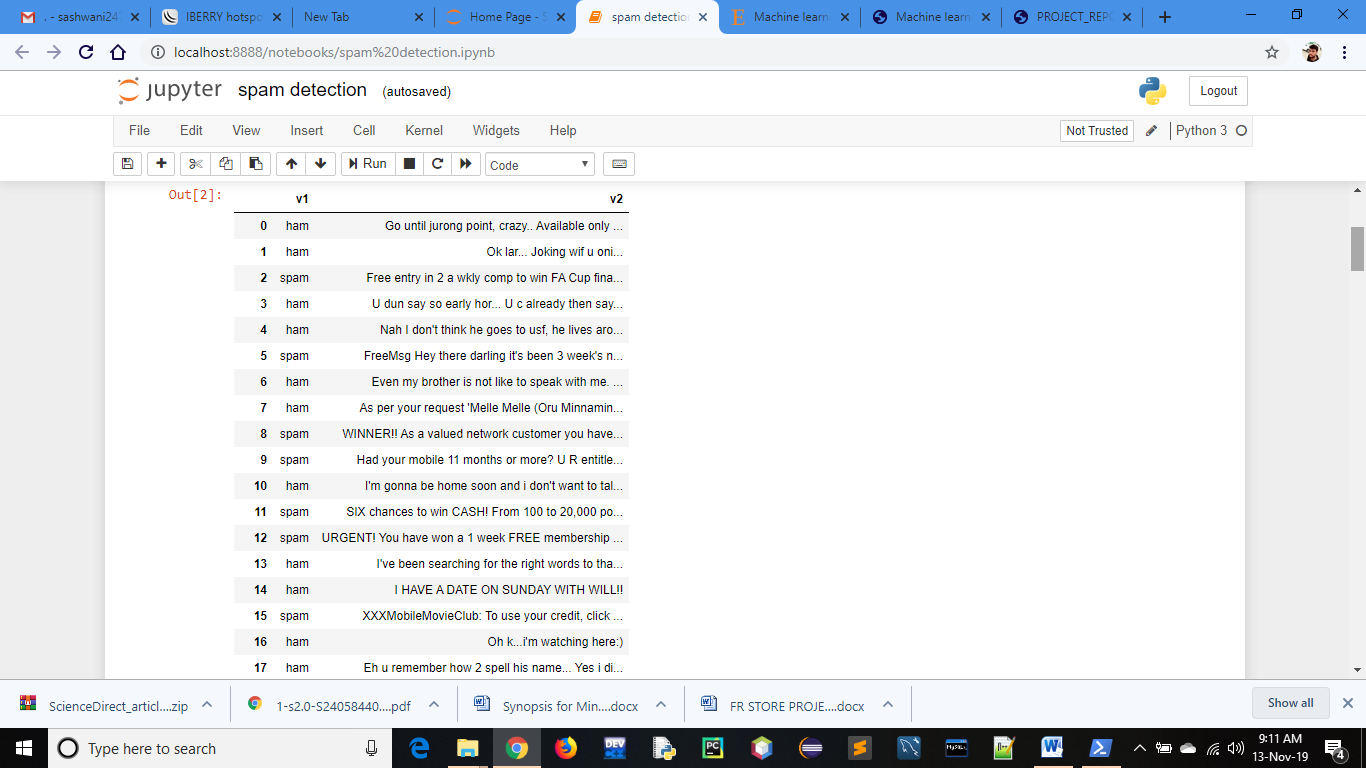
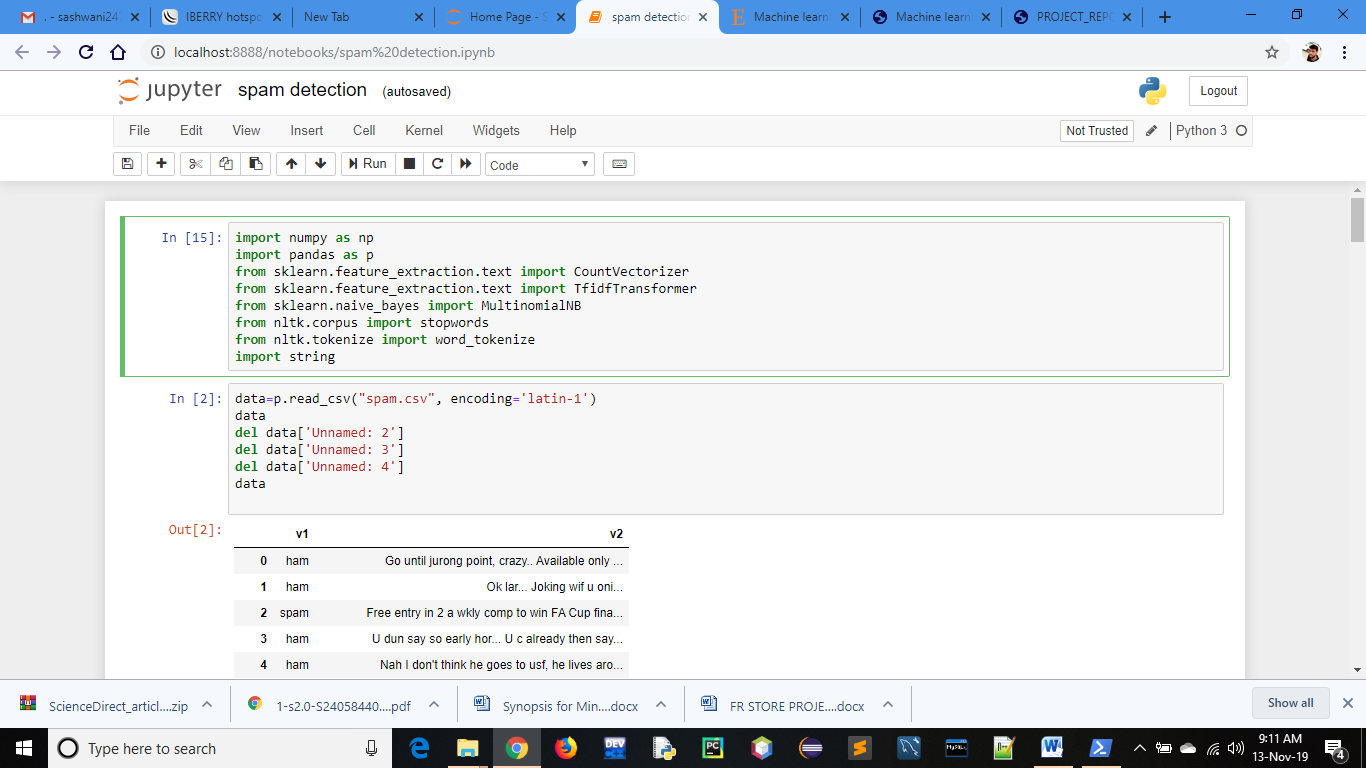
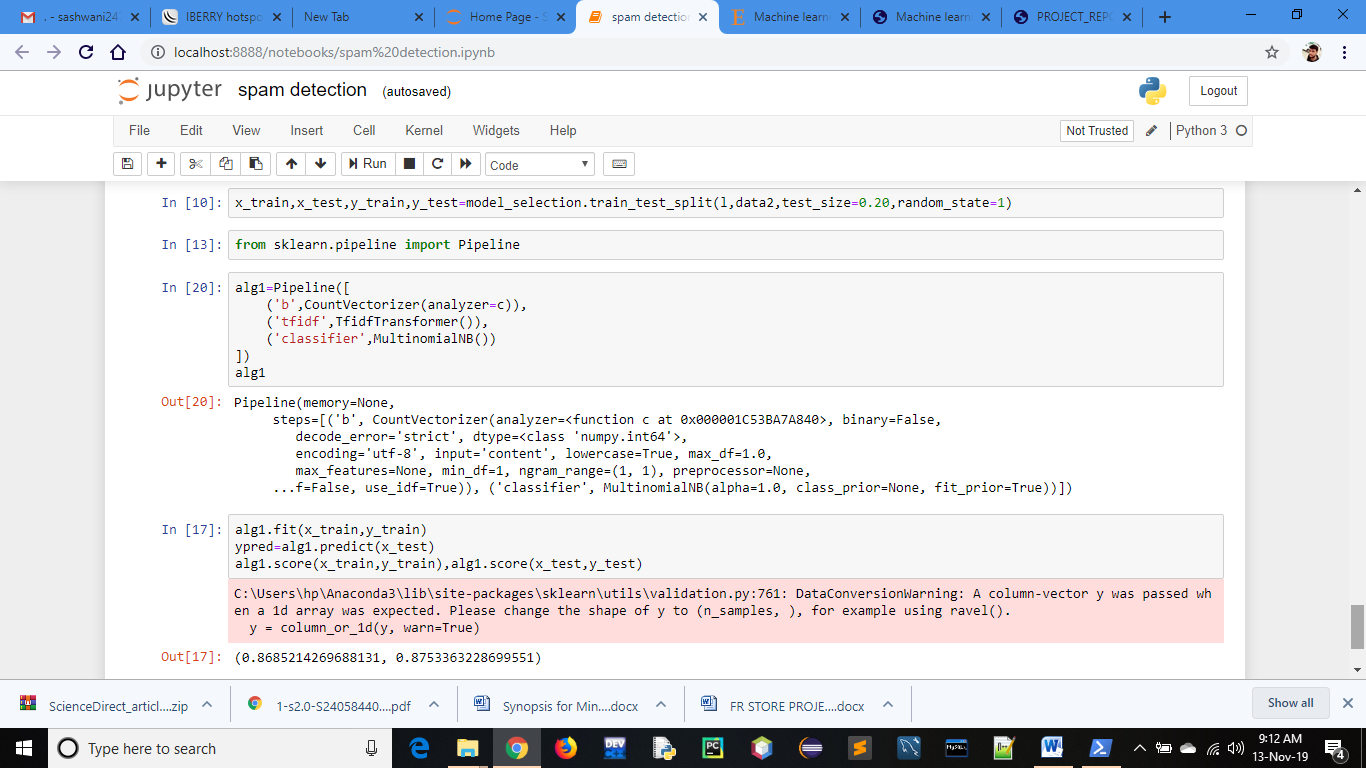
Future Scope:

1. Increasing the value of the content in Email
2. To Detect The Fraud Mails From Unknown Sources
3. We’ll understand the concepts, and we’ll have the knowledge.
4. To Create More Filtered Environment For More Safer and Secure Internet.

Chapter 5

**Expected Result :**

Screenshots for the Progress:



**Books / Sites referred :**

[http://en.wikipedia.org/wiki/ Machine](http://en.wikipedia.org/wiki/%20Machine) Leaning   
http://www.edx.org  
http://www.google.com