MINI PROJECT – II
(2018-19)

SPAM DETECTION

SYNOPSIS

Institute of Engineering & Technology

Team Members

Gaurav Bhardwaj
(161500215)
Mukul Bhardwaj
(161500331)
Deepak Singh
(161500189)
Kankshit Adhaulia
(161500254)

Supervised By
Subhash Agrawal
Asst. Professor
Department of Computer Engineering & Applications

About the Project:

What is SPAM? - One of the definition could be, electronic junk mail or junk newsgroup postings. Some people define spam even more generally as any unsolicited e-mail.
A spam filter is a software tool used to classify spam emails from genuine emails.Hence the spam filter predicts which class the email belongs to spam/no spam.When we choose to approach spam filtering from a machine learning perspective, we view the problem as a classification problem. That is, we aim to classify an email as spam or not spam depending on its features.We aim to study current methods in machine learning to identify the best techniques to use in spam filtering. We found that the Logistic Regression achieved the best performance.Words that frequently occur in a spam email are used as the feature set in the regression problem**.**

Motivation:

Email has become one of the most important forms of communication. In 2014, there are estimated to be 4.1 billion email accounts worldwide, and about 196 billion emails are sent each day worldwide.Spam is one of the major threats posed to email users. In 2013, 69.6% of all email flows were spam.Links in spam emails may lead to users to websites with malware which can access and disrupt the receiver's computer system.

Requirements:

Hardware:

- 4 GB RAM
- Window 10
- Disk Space : 4GB

Software:

- Anaconda
- Python