# ARIGNAR ANNA GOVERNMENT ARTS &

# SCIENCE COLLEGE FOR WOMEN

DEPARTMENT OF COMPUTER APPLICATION

# Project Name:

**Optimized Spam Filtering With Machine Learning**

Submitted By:

DIVYA.P

GOWSALYA.S

HARIPRIYA.S

HEMALATHA.S

# 

**OPTIMIZING SPAM FILTERING WITH**

**MACHINE LEARNING**

**INTRODUCTION:**

* Overview
* Purpose

**Overview :**

Spam-reduction techniques have developed rapidly over the last few years , as spam volumes have increased . We believe that the spam problem requires a multi-faced solution that combines a broad array of filtering techniques with various infrastructural changes ,changes in financial incentives for spammers ,legal approaches ,and more .This paper describes one part of a more comprehensive spam research .E-mail or Electronic mail is an electronic messaging system that transmits messages across computer networks .E-mail allow users to communicate with each other at a low cost as well as provides an efficient mail delivery system .The reliability ,user friendliness and availability of a wide range of free e-mail services make it most popular and a preferred communication tool .As such ,businesses and individual users alike heavily on this communication tool to share information and knowledge .Spam refers to bulk unsolicited commercial e-mail sent indiscriminately to users .Spam can be categorized into the following:

* Health
* Promotional products
* Financial and refinancing
* Phishing and other fraud
* Malware and viruses
* Education
* Marketing
* Political

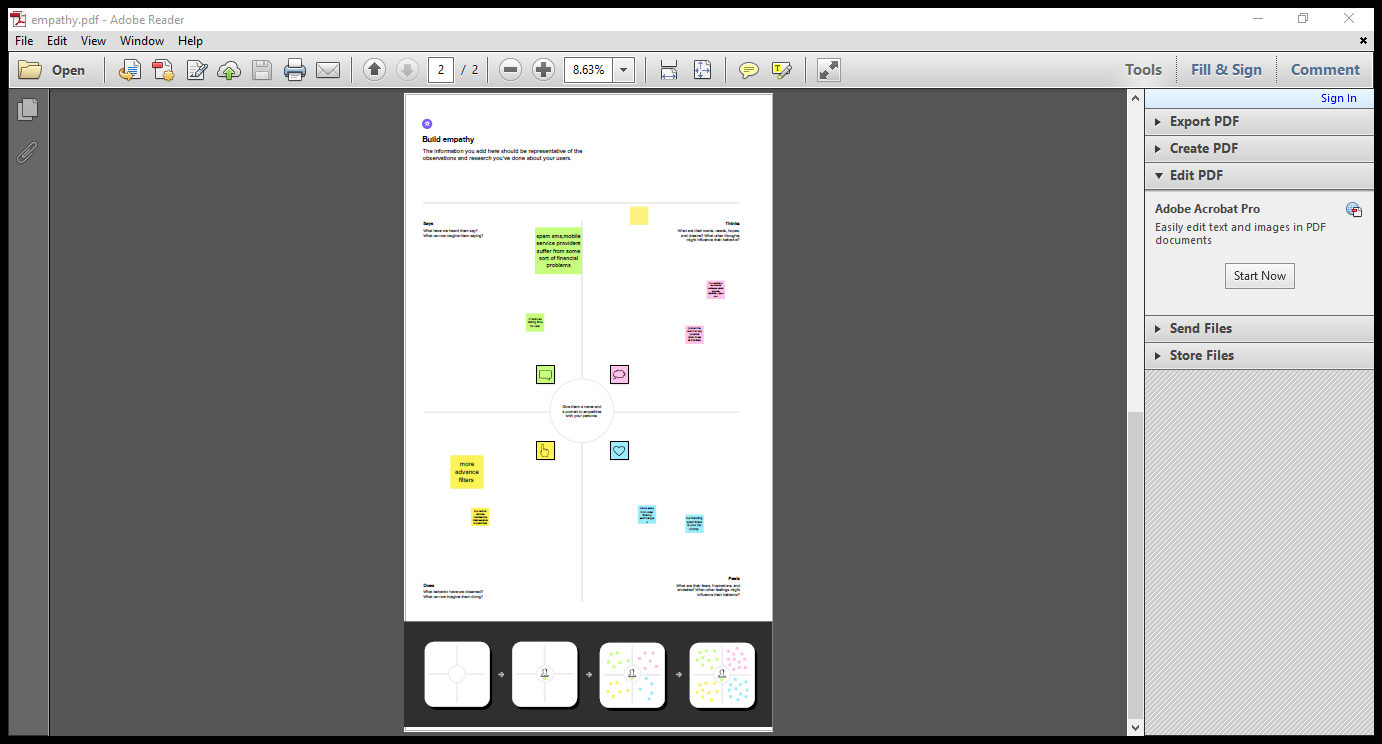
**Purpose:**

* Advertisements
* Pyramid schemes(Multi-Level Marketing)
* Giveaways
* Chain letters
* Political email
* Stock market advice

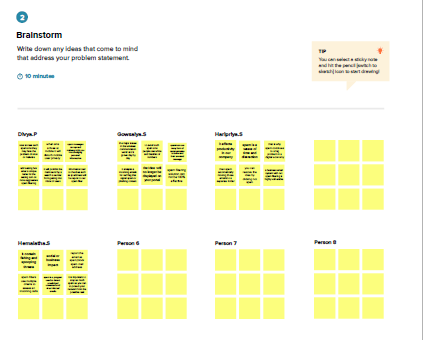
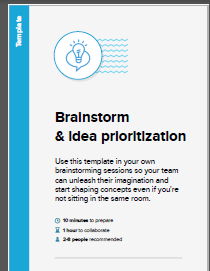
**Problem Definition and Design Thinking**:

1. Empathy Map
2. Ideation & Brainstorming Map

**Empathy Map:**

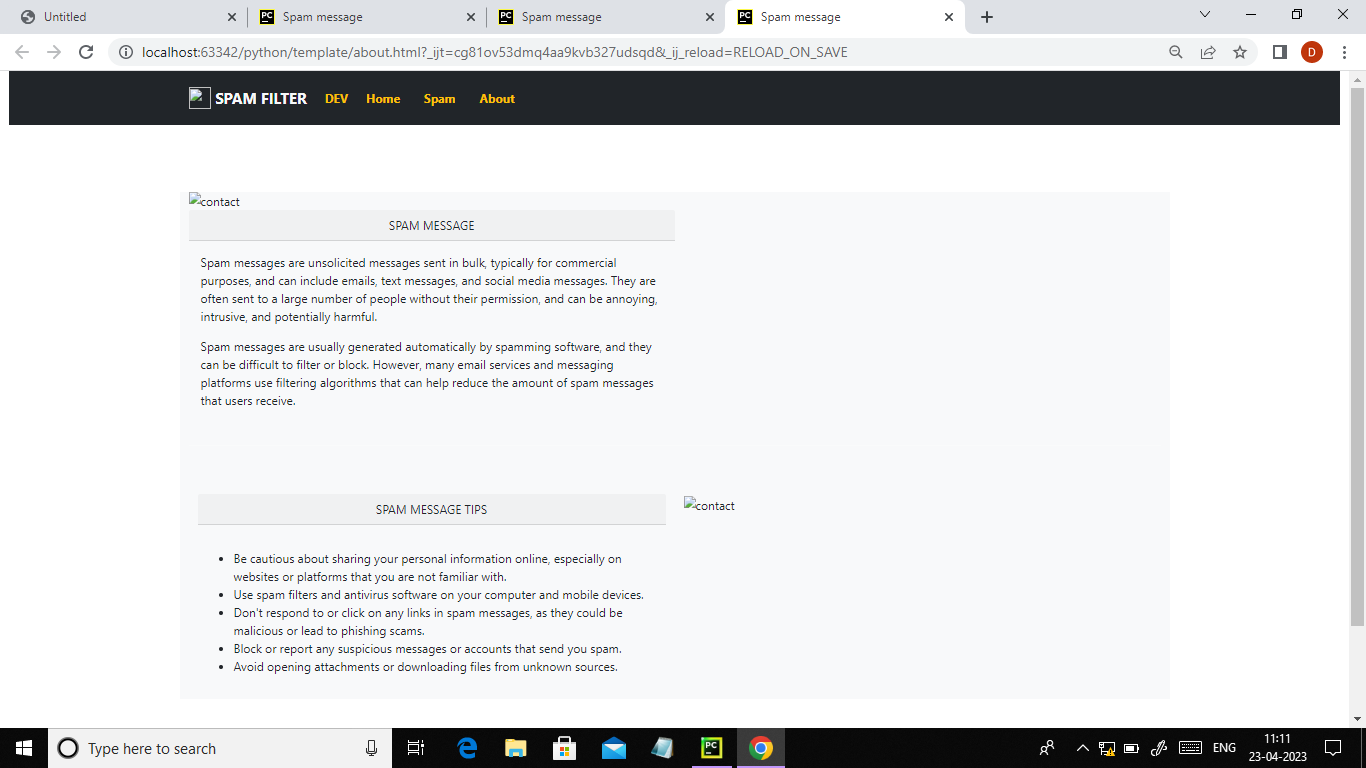


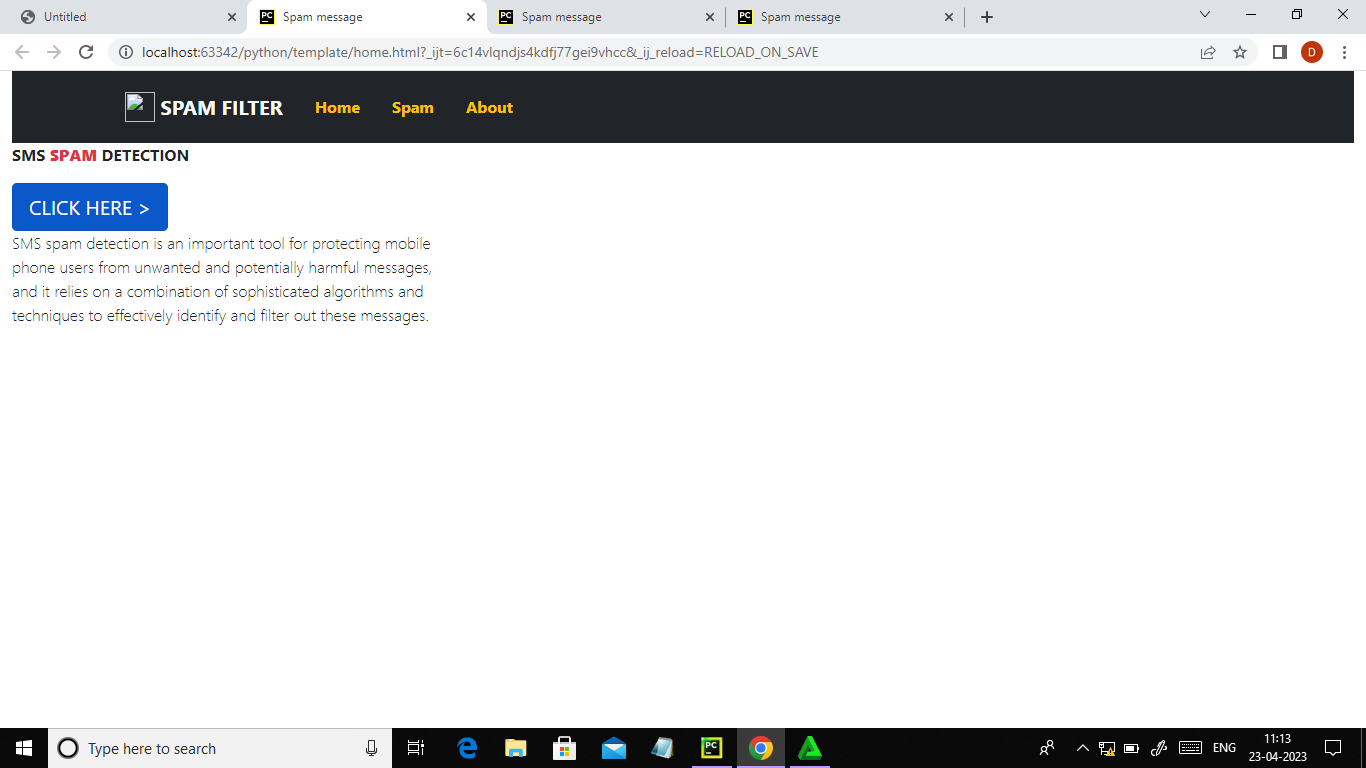
**Ideation & Brainstorming Map:**



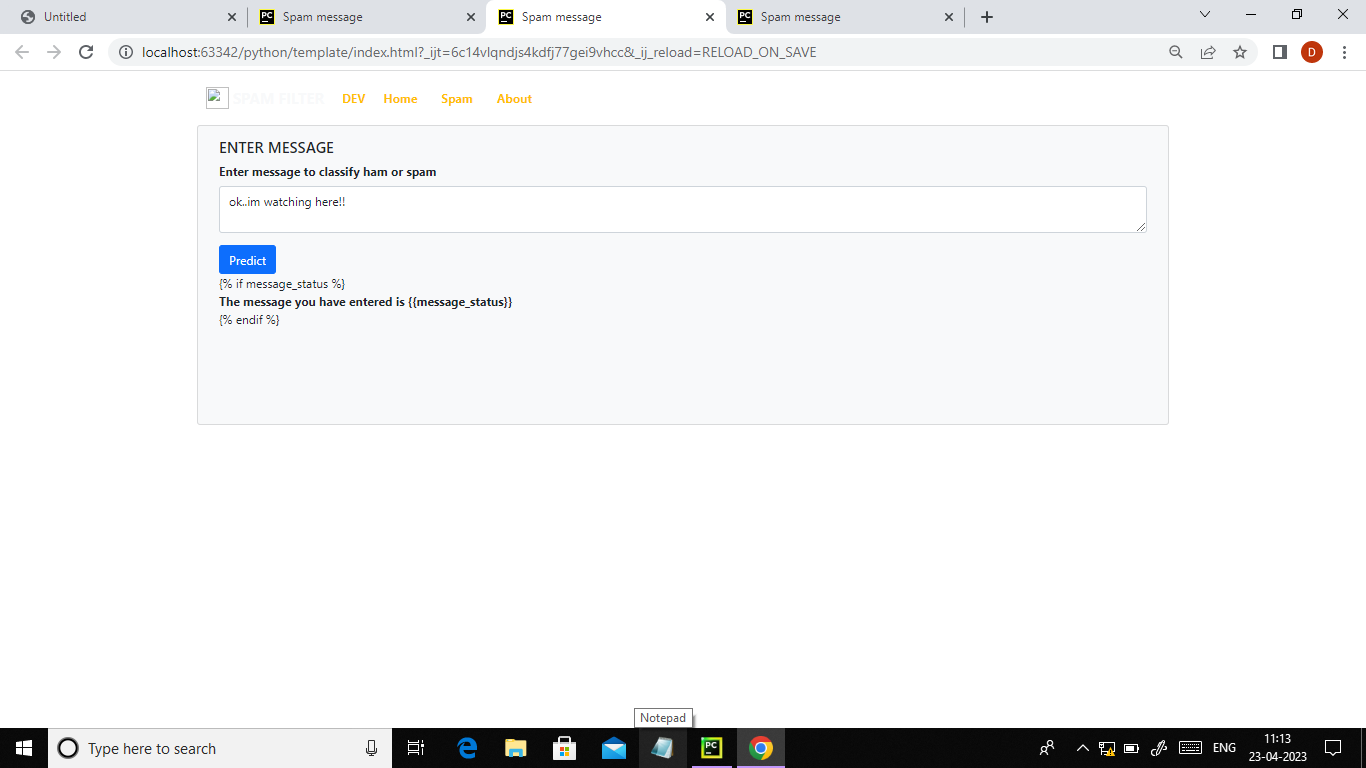
**RESULTS:**

About.html:



**Home.html:**

**Index.html:**

****

**App.py:**

****

**ADVANTAGES:**

* It streamlines inboxes
* Protect against malware
* Keeps you complaint
* It saves you money

**DISADVANTAGES:**

* Thousands of spam emails may reaches inboxes before a spammers ' s email address ,IP or domain is blacklisted.
* Spam filtering is machine based so there is a room for mistakes called "false positives".
* Bayesian filters may be fooled by spammers ,e.g.in a case of using large of legitimate text.

**APPLICATION:**

SMS Spam filters are typically implemented through software applications or services that analyze incoming messages and determine whether they are spam or legitimate. Some of the common applications used in SMS spam filtering include

* Machine Learning: Machine Learning algorithms can be trained on large datasets of messages to identify patterns and characteristics of spam messages. The algorithms can then use this knowledge to classify incoming messages as spam or legitimate.

**CONCLUSION:**

Email spam has been the focus of studies for a long time Though there are many different techniques to block spam email messages to reach users inbox ,filtering is the most commonly used mechanism and has gained success to some extent .Given the large number of usage of email worldwide email spam is still plentiful and scale of the problem is enormous .Researchers and organization make the filters smart and self-learning but spammers are stephead .They keep on finding techniques to deceive the filters and their learning mechanism .

Hence ,the problem still remains giving scope for researchers to work in the area .This work is an effort in the same scope to reduce false negatives/spam in the inbox of the users which has deceived the organizational filters .spam is becoming one of the most annoying and malicious addition to Internet Technology .Machine Learning approaches have provided researchers with a better way to combat spam .It has successfully applied in text classification .Since e-mail contain text ,the ML approach can be seamlessly applied to classified spam.

**FUTURE SCOPE:**

This work proposes a model for improving recognition of cruel spam in email. Our model resolve employ a novel dataset intended for the process of feature choice, and then validate the set of chosen features using three classifiers identified in spam detection: support,

Vector machine

Naive bayes

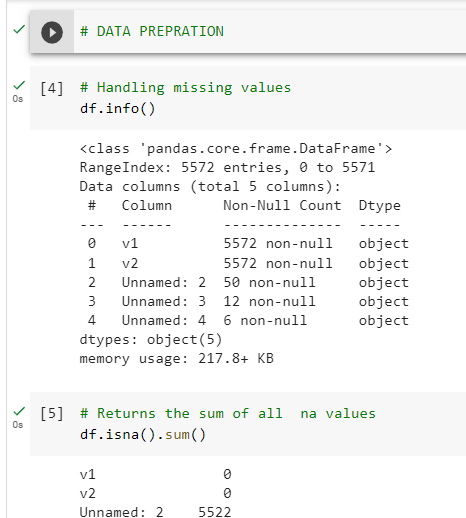
Multilayer perception

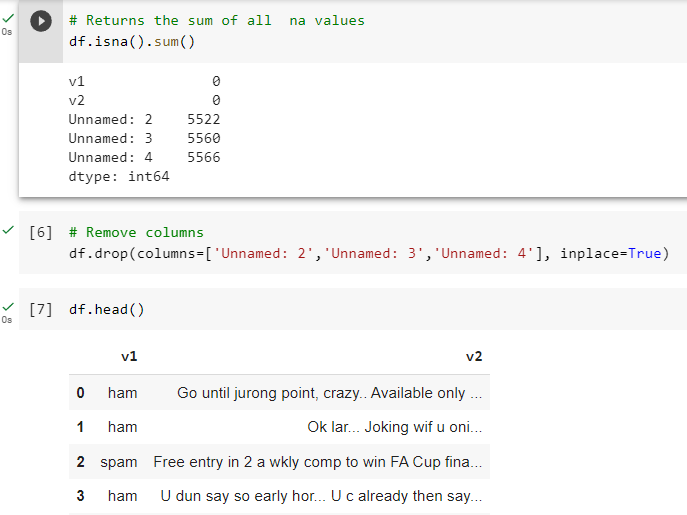
Feature selection is projected to recover training time as well as accuracy for the classifiers.

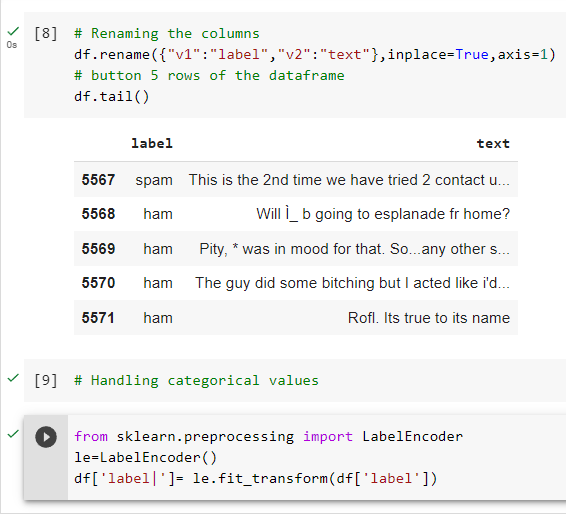
**APPENDIX:**

Source code:

****







\

