

Bharat Institute of Engineering and Technology (BIET)

Mangalpally (V), Ibrahimpatnam (M), Ranga Reddy (D), Hyderabad

2019 – 2020 B. Tech. Final Year Major Project

Department of Computer Science Engineering

Data security approach on cyber-crime with web vulnerability

Batch Number : 16CSC16IMP-04

16E11A05F6 AKHIL GUPTA

16E11A05D3 ABDUL MUQEET OSAMA

16E11A05E7 K VISHAL PATEL



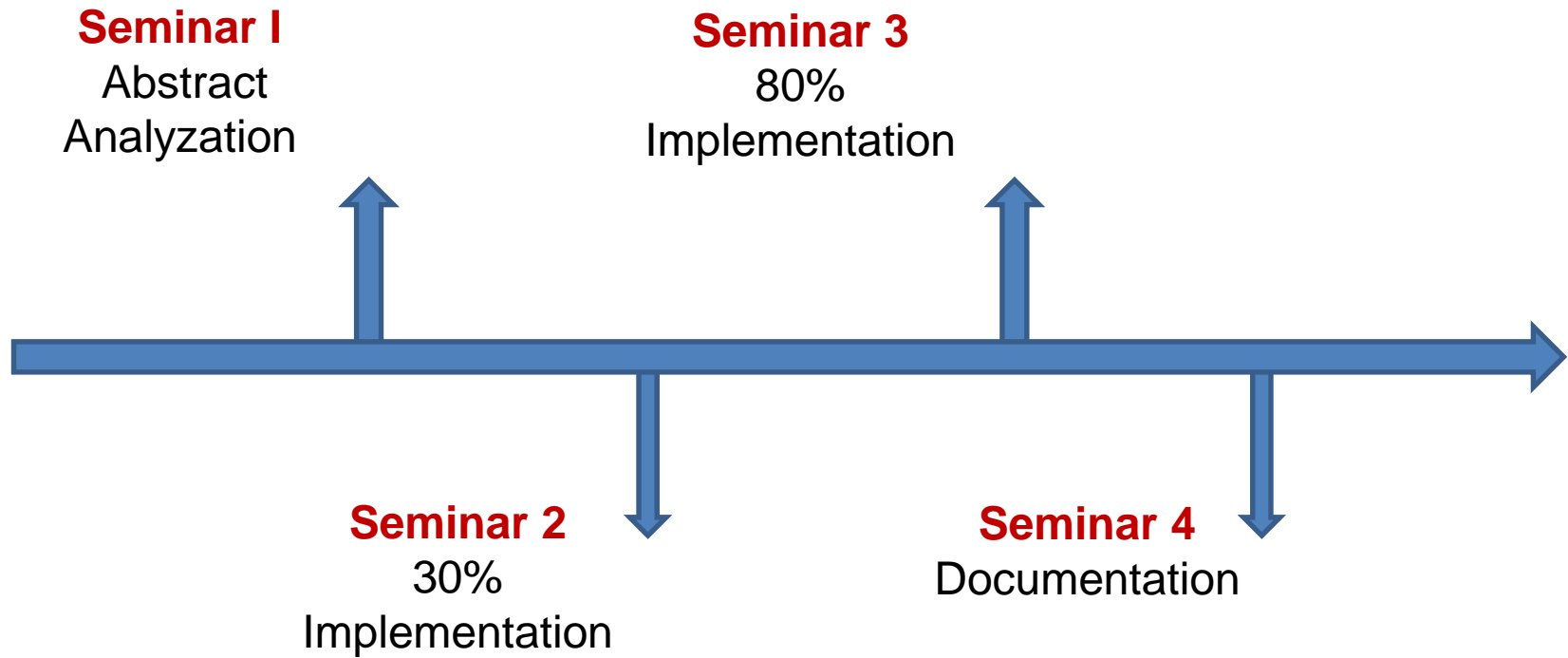
**BHARAT
INSTITUTIONS**
Imparting Value Based Education

Our Specific Contribution

Student ID	Student Name	Contribution
16E11A05F6	Akhil Gupta	Implementation of Controllers and Templates
16E11A05D3	Abdul Muqeet Osama	Implementation of Models and Views
16E11A05E7	K Vishal Patel	Implementation of Designing ,Use Case, Class, Sequence and Activity Diagram.



Project Timeline



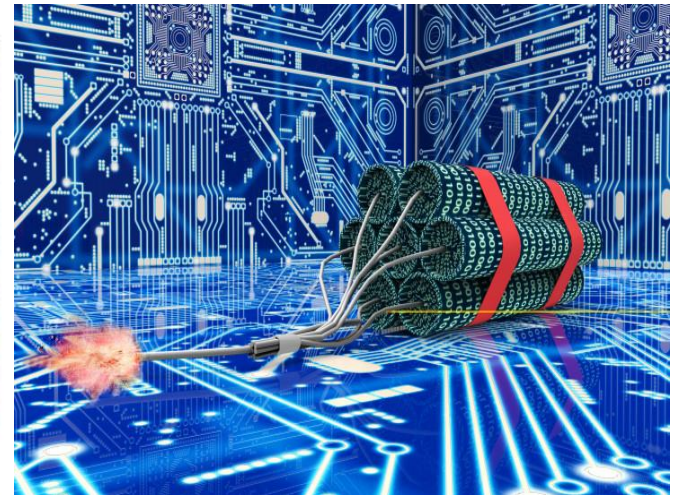
Our Base/Reference Paper

- “Analysis of Text Mining Techniques over Public Pages of Facebook.”
- This was published in IEEE 6th International Conference on Advanced Computing.
- Year: 2016
- This is an conference.
- Yes, This is a Scopus paper.



Introduction

TERRORISM, THE INTERNET AND SOCIAL MEDIA-



THE INTERSECTION OF TERROR AND TECHNOLOGY

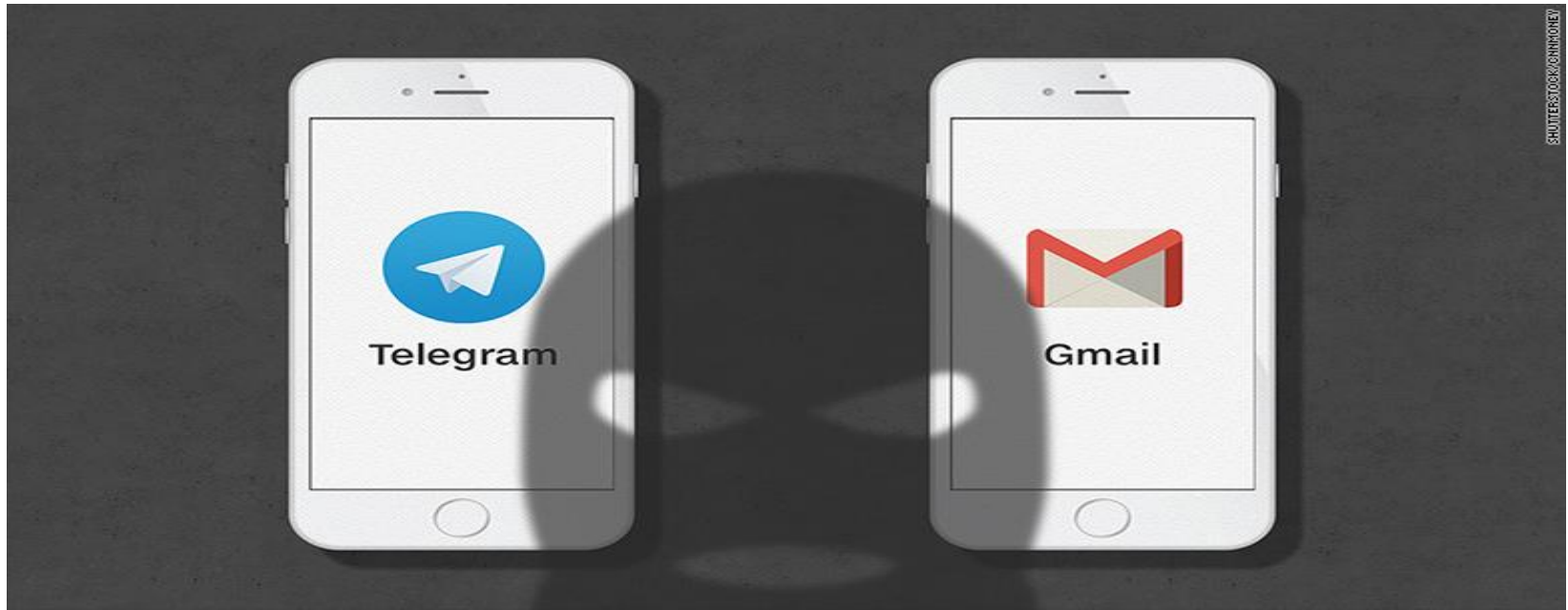


**BHARAT
INSTITUTIONS**
Imparting Value Based Education

WHY THE INTERNET IS AN IDEAL TOOL FOR TERRORIST ORGANISATIONS

- Enables rapid communications.
 - Ease of access.
 - Lack of regulation.
 - Fast flow of info.
 - Vast potential audiences.
- Low cost.
- Ubiquity of the internet.
- Growth in band width - complex info.
- Anonymity by encryption - offers a 'virtual sanctuary'.





<https://money.cnn.com/2016/05/03/technology/jihadist-tech-tools/index.html>

- The free, mobile chatting service Telegram -- which is anonymous and encrypted -- is used by 34% of jihadists whose communications were the subject of a [study](#) by cybersecurity firm Trend Micro ([TMICF](#)).
- Gmail is the standard email service used by 34% of jihadists, according to Trend Micro. This shows terrorists have simply gravitated toward the most popular email service in the world, which now has 1 billion users.

Bomb Threat Emails: Extortion Gets Physical

From [redacted] ☆

Subject **Bomb is in your building**

To [redacted] ☆

↩ Reply ↩ Reply All ▾ ➔ Forward More ▾

Good day. My recruited person hid the bomb (lead azide) in the building where your business is located. It was built according to my instructions. It has small dimensions and it is covered up very carefully, it can not damage the supporting building structure, but if it explodes there will be many victims. My mercenary keeps the territory under the control. If he notices any unnatural activity or cops he will power the device.

I can withdraw my mercenary if you make a transfer. You send me 20.000 \$ in BTC and the device will not explode, but don't try to fool me -I ensure you that I will withdraw my recruited person solely after 3 confirmations in blockchain.

Here is my btc address : 1BHasGex1jhrZeY7KyUGGKUNRtVgKedRY8

You must transfer bitcoins by the end of the workday, if you are late with the payment the device will detonate.

This is just a business, if you don't transfer me the money and a bomb detonates, other commercial enterprises will pay me more money, because this isnt a one-time action.

To stay anonymous I will not visit this email account. I monitor my wallet every 30 minutes and if I see the money I will give the command to my recruited person to get away.

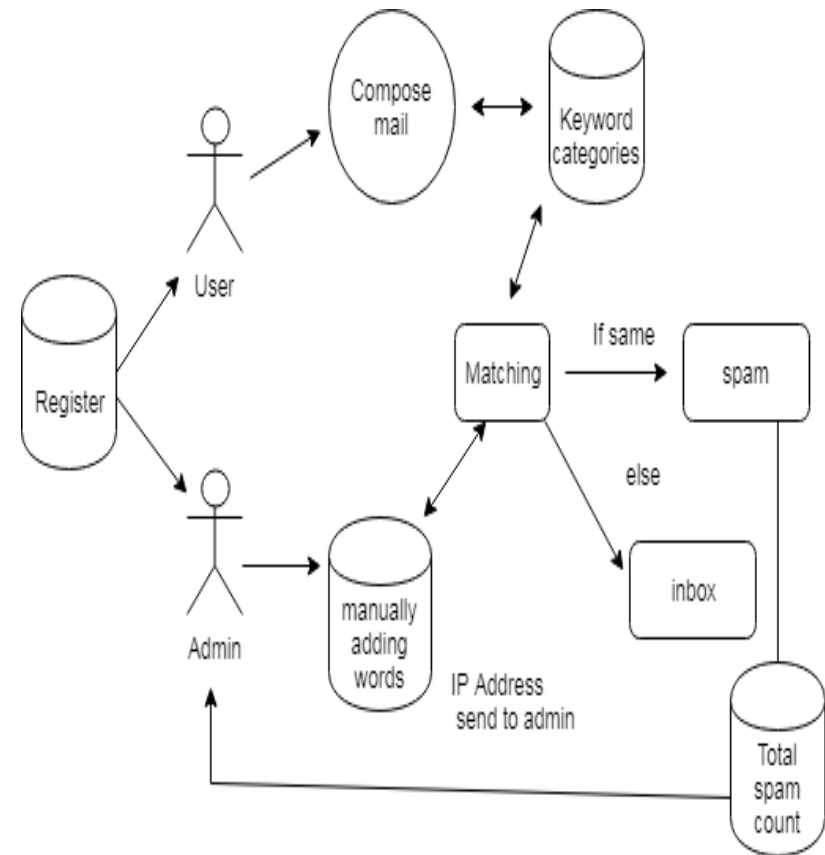
If the bomb blows up and the authorities see this letter-

We are not terrorists and do not take liability for explosions in other buildings.

<https://www.digitalshadows.com/blog-and-research/bomb-threat-emails-extortion-gets-physical/>



- We propose an efficient web Data Mining System.
- We utilized Data Mining in Email System to identify the undesirable messages that are progressively helpless to terrorism.
- These will send to the spam legitimately to the beneficiary who is utilizing the System.



Related Work/Literature Review 1

- “Analysis of Text Mining Techniques over Public Pages of Facebook “.
- Applying web mining Techniques in spam detection of emails.
- Disadvantages of base paper:
Spam activities are not limited to one domain of Web.
It is not possible to remove spam completely from the sphere of the web.



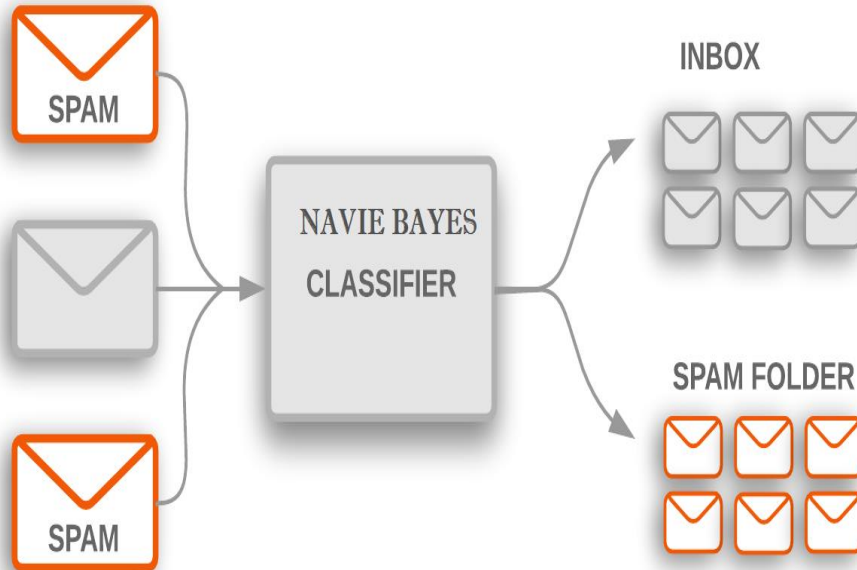
Related Work/Literature Review 2

- To overcome these disadvantages, we need acquire more knowledge about web mining.
- And text mining then followed by the Classification Technique to increase the accuracy of the Spam detection activity.
- Novelty of Email Spam Detection for Terrorism related data.



Our Approach

- Mailing
- Filtering
- Spam Detection
- Preprocessing
- Algorithm



Mailing Module

- User Registration and Login.
- It is E-mail system. It contain the highlights of inbox, sent mail, spam, recent histories, and so on.,
- The client can make the mail to sender. It might be identified with terrorism or non terrorism.
- Recent history signifies the individual who is doing this mail.



Filtering Module -1

- We have a few data's in our dataset.
- This will check whether sent message have contained the filtration words about terrorism or not?
- By Utilizing Data mining strategy to mine out text data from large datasets and utilize get the results.



Filtering Module -2

- Web mining comprises of text mining systems. Through that text mining, we can extract the text or content what are totally identified with psychological warfare.
- The filtration words are coordinate with the sent message implies, the beneficiary gets the mail in his/her spam box or else inbox.



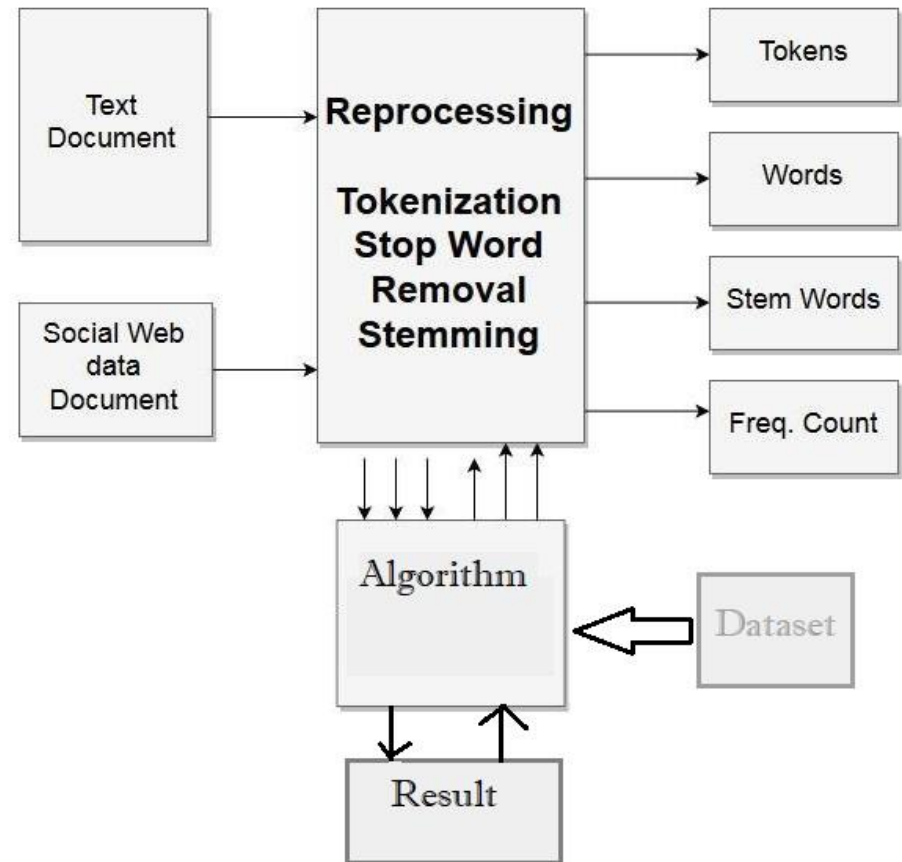
Spam Detection Module

- Admin login.
- Admin will have the highlights of keywords, spam, investigation, diagram.
- By utilizing Mining ideas Admin can include not many terrorisms related words.
- In examination, it contains a mail having what number of words in those keyword categories and their total count per each mail.



Preprocessing Module

- Spam Detection will contain preprocessing which implies it will expel all the normal words/stop words, for example:
the, and, or, here, there, and so on.,
- Here we have utilized the Naïve Bayes calculation.



Algorithm



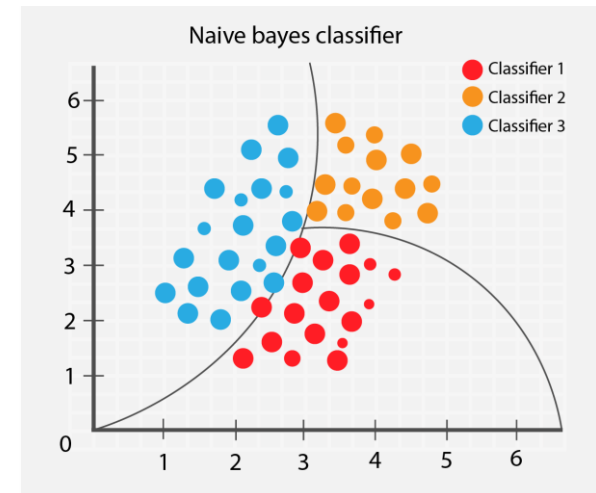
Naive Bayes Algorithm

- A naive Bayes classifier is an algorithm that uses Bayes' theorem to classify objects.
- Naive Bayes Algorithm is based on Bayes' theorem with the assumption of independence between every pairs of features.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

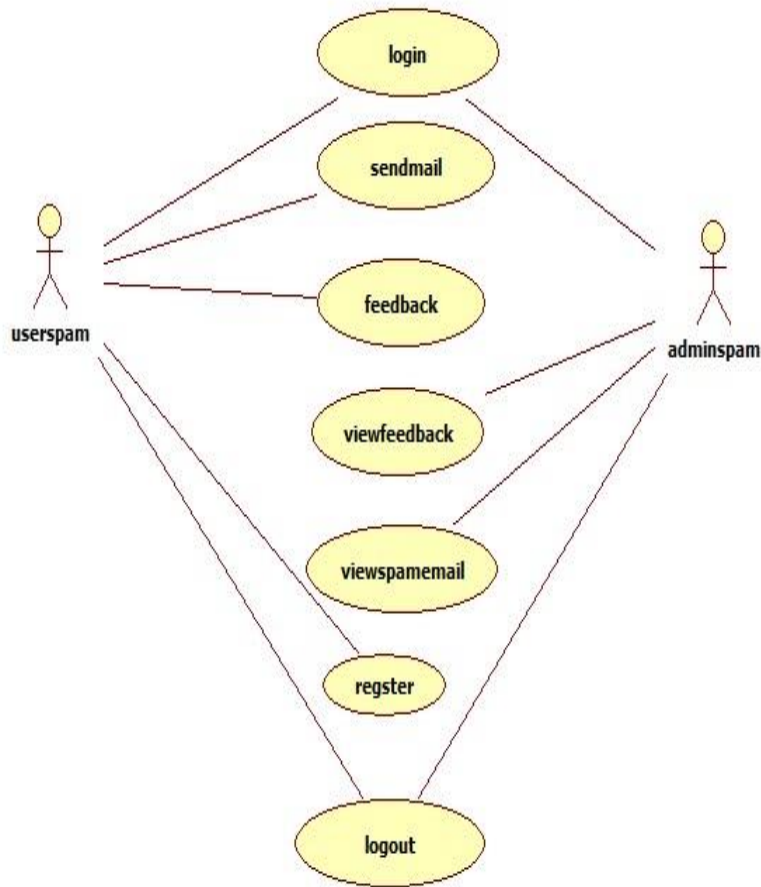
using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

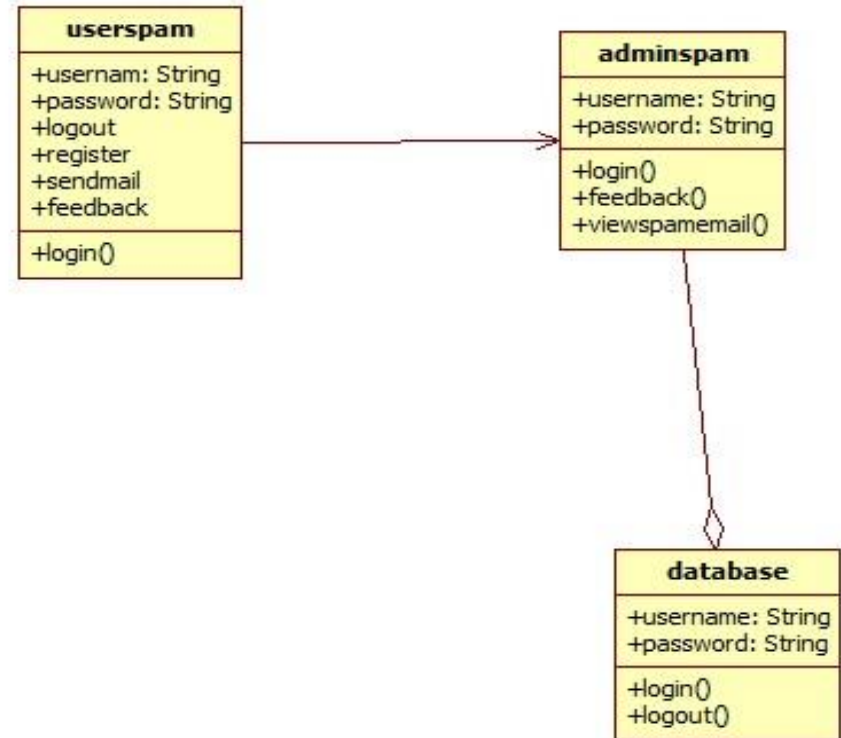


Flow Graphs

Use cases Diagram

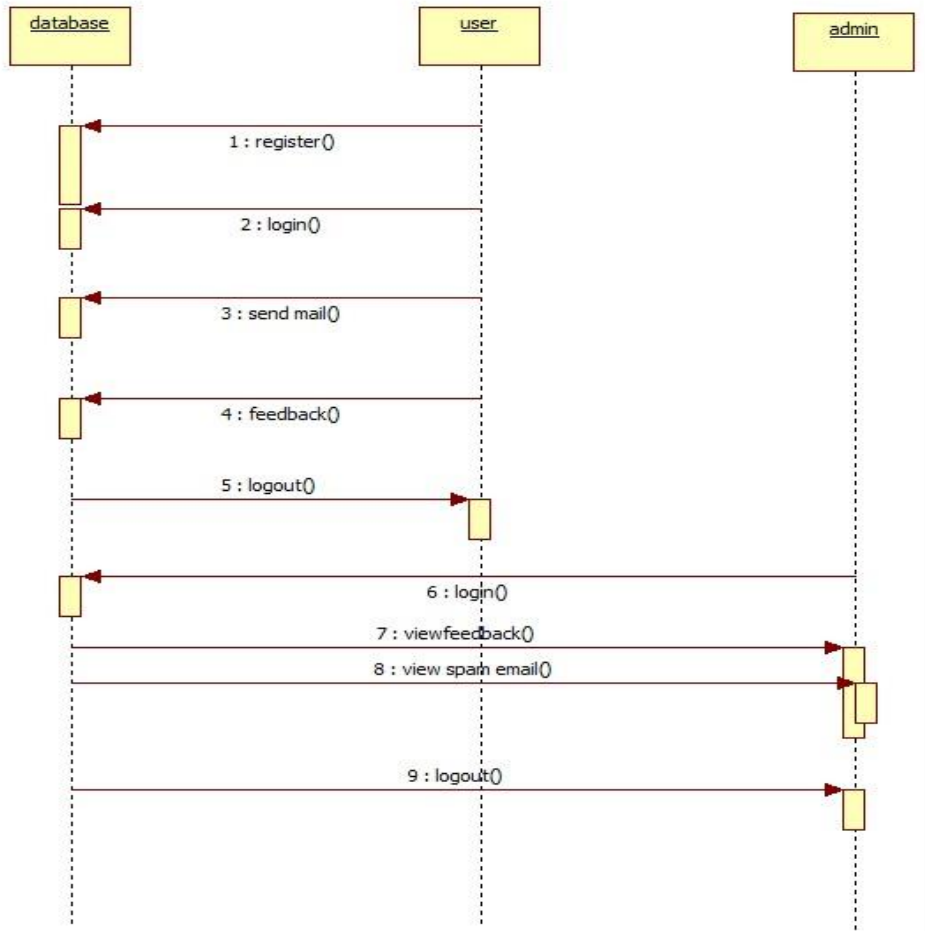


Class Diagram

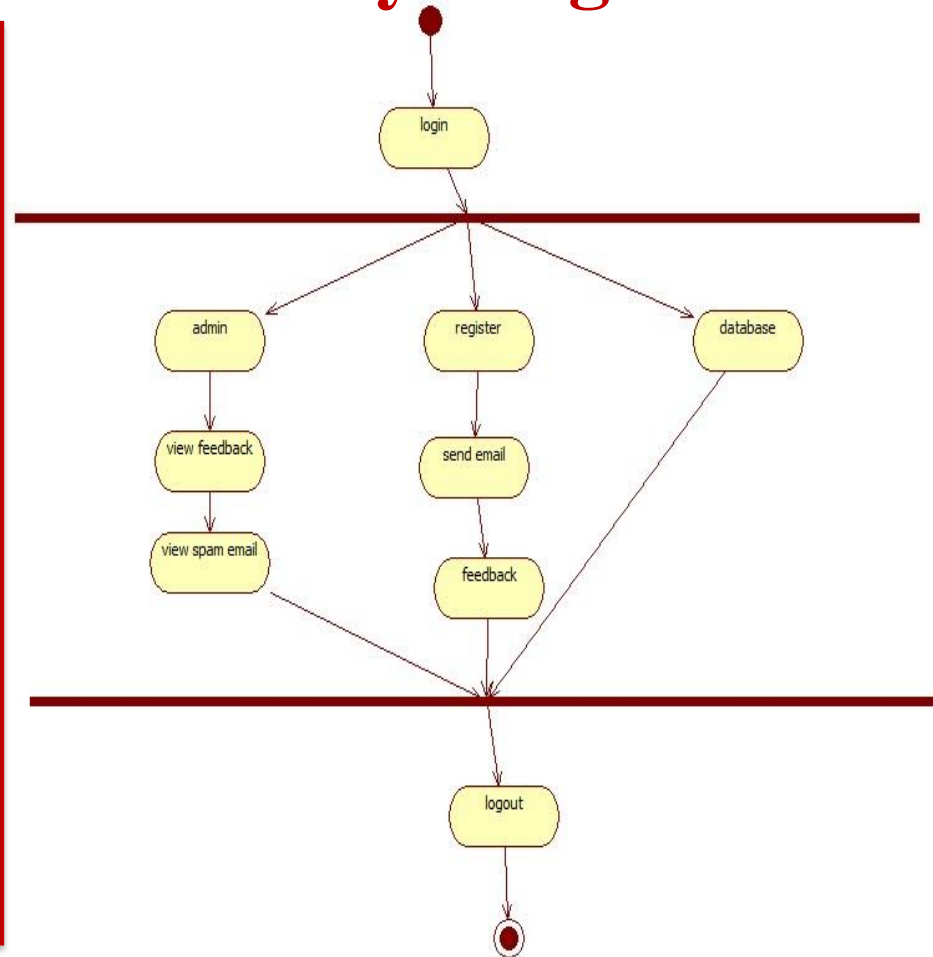


Flow Graphs - 1

Sequence Diagram



Activity Diagram



Justifying Our Approach -1

- A naive Bayes classifier is an algorithm that uses Bayes' theorem to classify objects.
- Naive Bayes classifiers assume strong, or naive, independence between attributes of data points.
- This algorithm requires small amount of training data to estimate the necessary parameters.
- Naive Bayes classifiers are extremely fast compare to more sophisticated methods.



Justifying Our Approach -2

Accuracy :

$$\frac{\text{True Positive} + \text{True Negative}}{\text{Total Population}}$$

F1- Score :

Average Weight of
Precision and Recall

Classification Algorithms	Accuracy	F1-Score
Logistic Regression	84.60%	0.6337
Naïve Bayes	80.11%	0.6005
Stochastic Gradient Descent	82.20%	0.5780
K-Nearest Neighbours	83.56%	0.5924
Decision Tree	84.23%	0.6308
Random Forest	84.33%	0.6275
Support Vector Machine	84.09%	0.6145



Justifying Our Approach -3

Disadvantages in classification algorithms:

Algorithm:

- Logistic Regression : Binary Prediction
- Naïve Bayes : Bad Estimator
- Stochastic Gradient Descent : Requires hyperparameters, features sensitive
- K- Nearest Neighbors : high computation cost, determine the K value
- Decision Tree : Unstable, complex tree will generate
- Random Forest : Slow real time prediction, difficult and complex implementation
- Support Vector Machine : No direct probability estimation



Experimental Setup

HARDWARE REQUIREMENTS:

- **System** : Pentium IV 2.4 GHz.
- **Hard Disk** : 40 GB.
- **Floppy Drive** : 1.44 Mb.
- **Monitor** : 14' Colour Monitor.
- **Mouse** : Optical Mouse.
- **Ram** : 512 Mb.

SOFTWARE REQUIREMENTS:

- **Operating system** : Windows 10.
- **Coding Language** : Python.
- **Framework** : Django.
- **Front-End** : Python.
- **Designing** : HTML5, CSS, JavaScript.
- **Data Base** : MySQL.



Experimental Setup

- Installing in Python



- Installing and Setup Virtual Environment:

```
pip install virtualenvwrapper-win  
mkvirtualenv mywebsite  
workon mywebsite
```

- Installing and Setup Django:

```
pip install Django
```



- Checking the Version of Django after Installation:

```
python -m django --version
```



Experimental Setup

Creating and Running a Django Project

- `cd \`
- `mkdir Major_project`
- `cd Major_project`
- `django-admin startproject email_spam_detection`
- `cd email_spam_detection`
- Running Server:
`python manage.py runserver`
- Setup database:
`python manage.py makemigration`
`python manage.py migrate`



Project Application

- Our project can be utilize in E-mail System to identify the undesirable messages
- Our project is to solve an real-time issue of undesirable messages or Spam messages which is received via E-mail .
- Cost of implementation of our project work in real world is less \$2000 USD for complete deployment of project Production Server
- Drawbacks of our project will be not able stop sent phony news or messages through the Social media Platform.



Conclusion

- To curb and destroy the terrorism and spreading of their activities .
- Through online social media through unwanted messages and images to cover the helpless people.
- We need to use the powerful method or system.
- This system should be useful to the cops for easily give awareness to common people.
- Find the person who are spreading the harmful words as well as who are all involved in terrorism.



Future Work

- Tracking and Tracing of Digital Evidence against the culprit for illegal activities.
- To track the encrypted mails and documents should be tracked by system. If any undesirable content found in the document need to be forwarded directly to the admin as alert message with the sender details.
- To apply this system on other social media platforms like what's app, telegram for unusual forwards.



Thank you



BHARAT
INSTITUTIONS
Imparting Value Based Education