**MACHINE LEARNING WITH PYTHON**

A Project Report Submitted to the Bharathidasan University in partial

fulfillment of requirement for the award of the degree of

**BACHELOR OF COMPUTER SCIENCE**

Under the guidance of

**Dr. K. ELANGOVAN, M.Sc., M.Phil., Ph.D**

Guest Lecturer

****

**DEPARTMENT OF COMPUTER SCIENCE**

**GOVERNMENT ARTS AND SCIENCE COLLEGE (Co-Ed),**

(Affiliated to Barathidasan University)

**THARAGAMPATTI-621 311.**

**APRIL -2023**

PROJECT REPORT

**OPTIMIZING SPAM FILTERING WITH**

**MACHINE LEARNING**

**Team Leader** - Gayathri R

**Team Member**  - Lakshmi N

## Team Member - Mathi selvi S

## Team Member - Muniyammal M

**OPTIMIZING SPAM FILTERING WITH MACHINE LEARNING**

**1.INTRODUCTION**

**1.1 Overview**

**A brief description about your project:**

Over recent years, as the popularity of mobile phone devices has increased, Short Message Service (SMS) has grown into a multi-billion dollar industry. At the same time, reduction in the cost of messaging services has resulted in growth in unsolicited commercial advertisements (spams) being sent to mobile phones. Due to Spam SMS, Mobile service providers suffer from some sort of financial problems as well as it reduces calling time for users. Unfortunately, if the user accesses such Spam SMS they may face the problem of virus or malware. When SMS arrives at mobile it will disturb mobile user privacy and concentration. It may lead to frustration for the user. So Spam SMS is one of the major issues in the wireless communication world and it grows day by day.

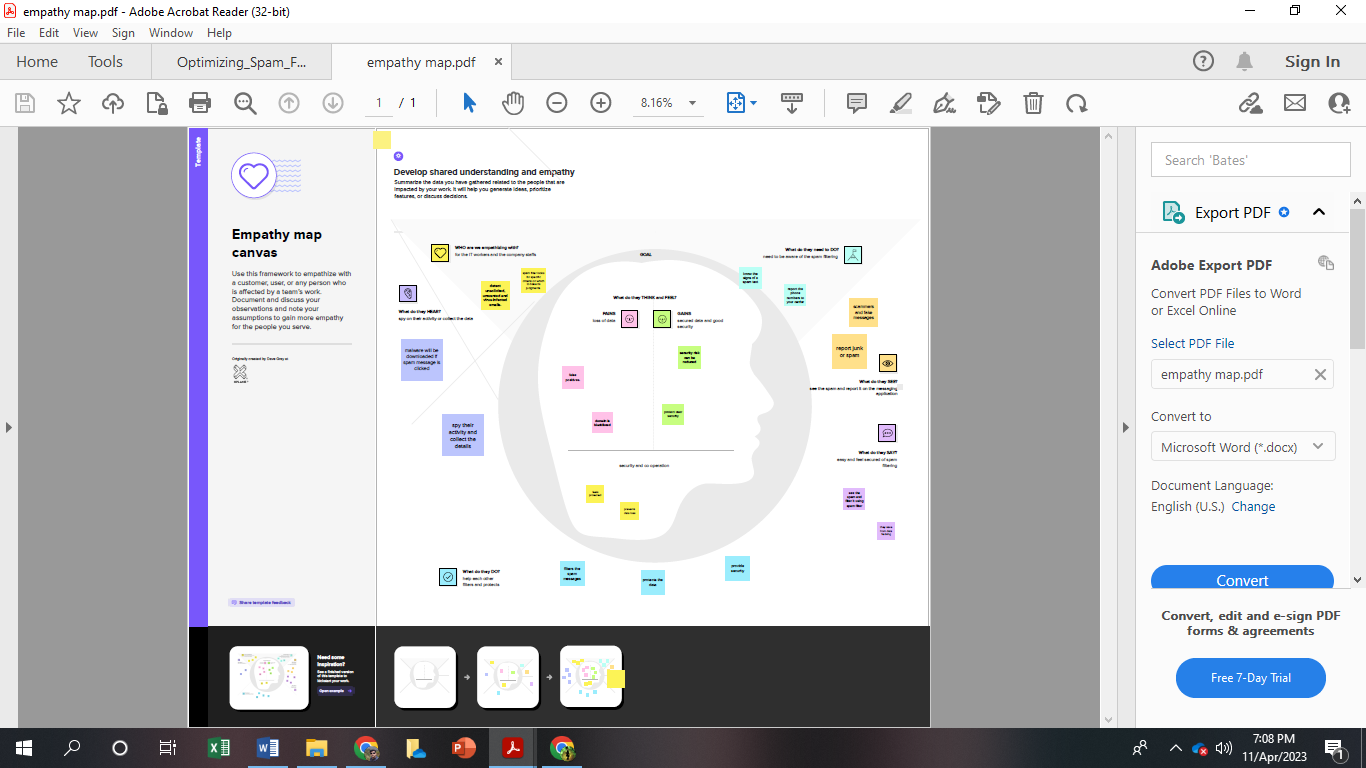
To avoid such Spam SMS people use white and black list of numbers. But this technique is not adequate to completely avoid Spam SMS. To tackle this problem it is needful to use a smarter technique which correctly identifies Spam SMS. Natural language processing technique is useful for Spam SMS identification. It analyses text content and finds patterns which are used to identify Spam and Non-Spam SMS.

**1.2 Purpose**

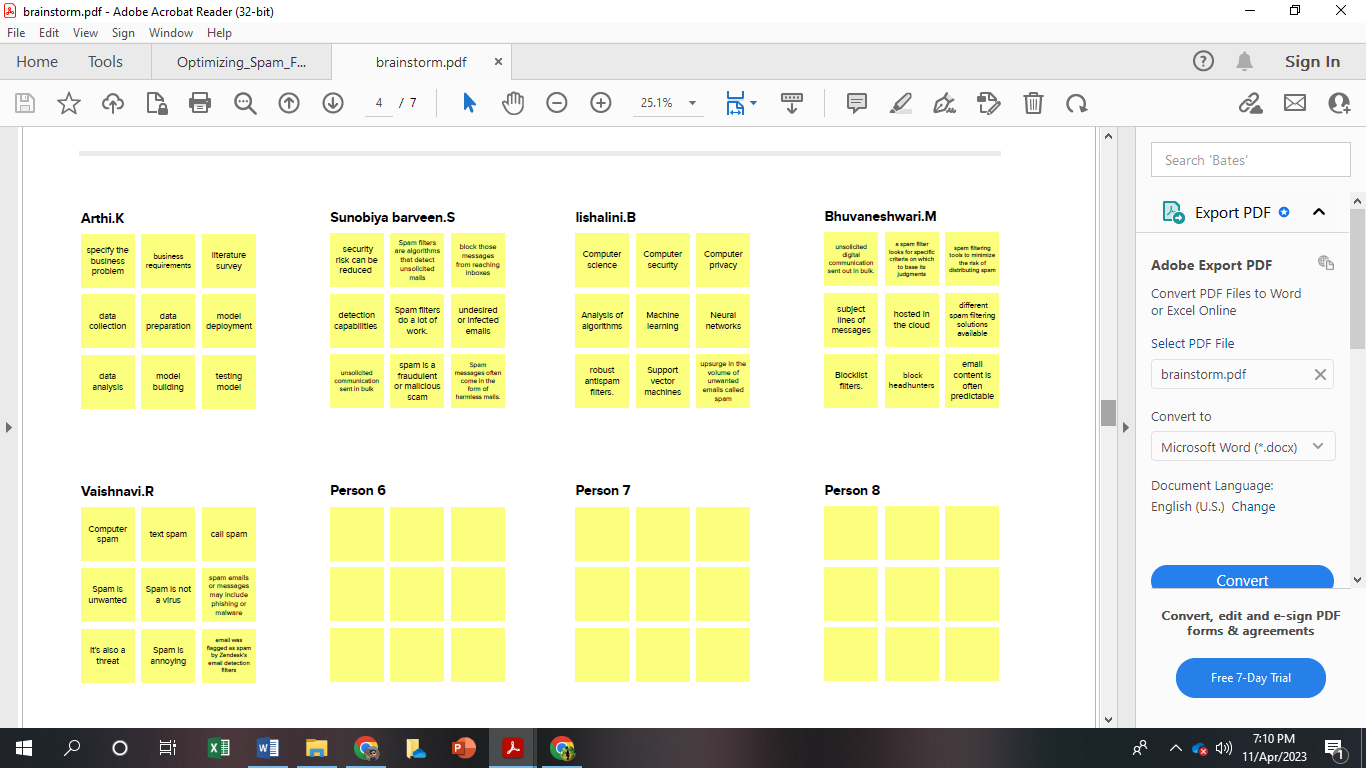
* User interacts with the UI to enter the input.
* Entered input is analysed by the model which is integrated.
* Once model analyses the input the prediction is showcased on the UI.

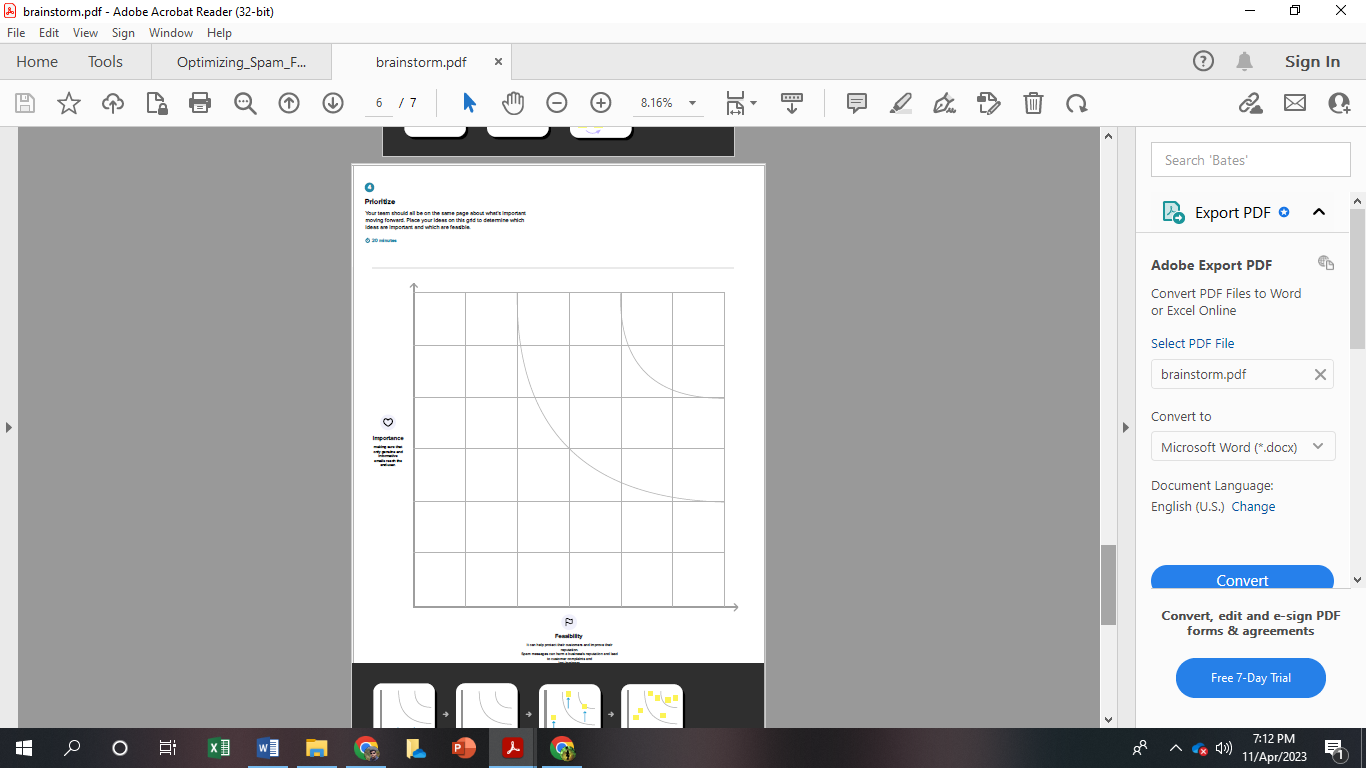
**2. DEFINE PROBLEM & DESIGN THINKING**

**2.1 Empathy Map**

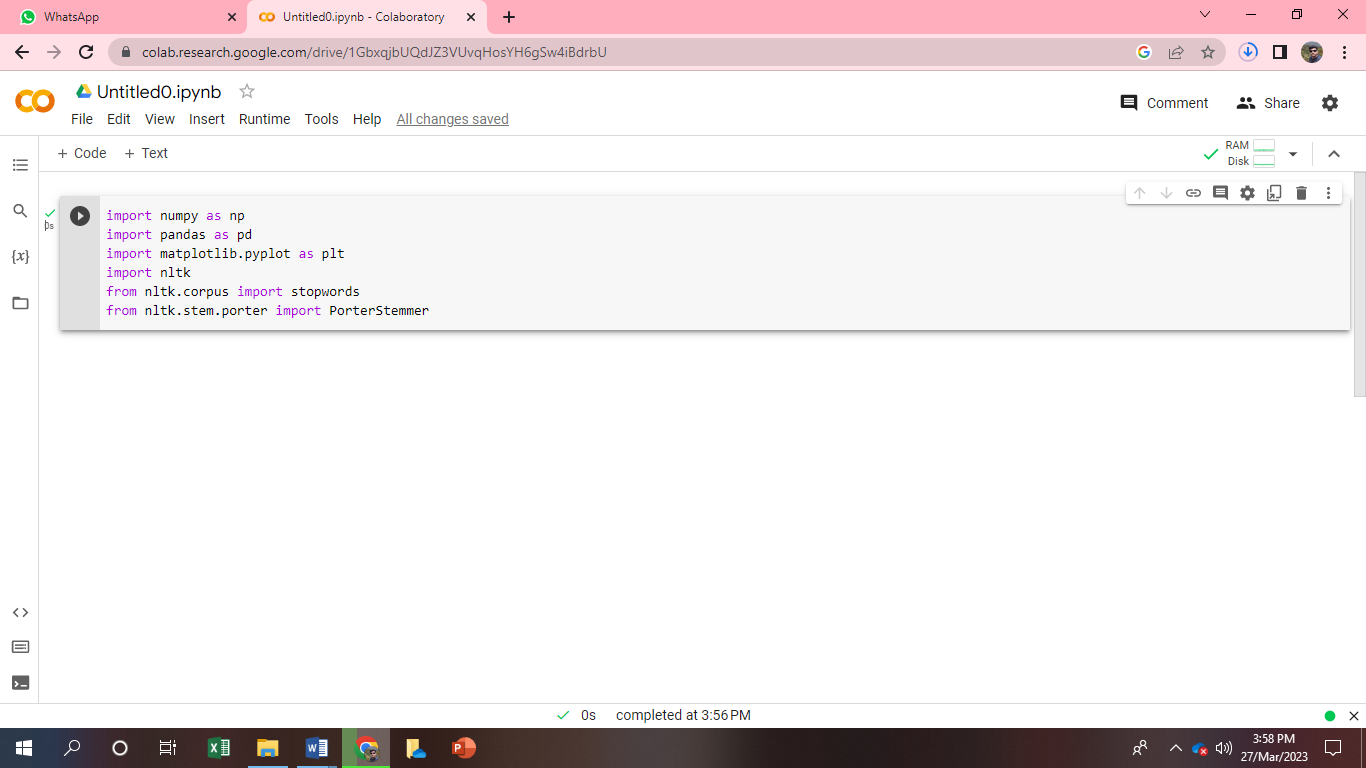


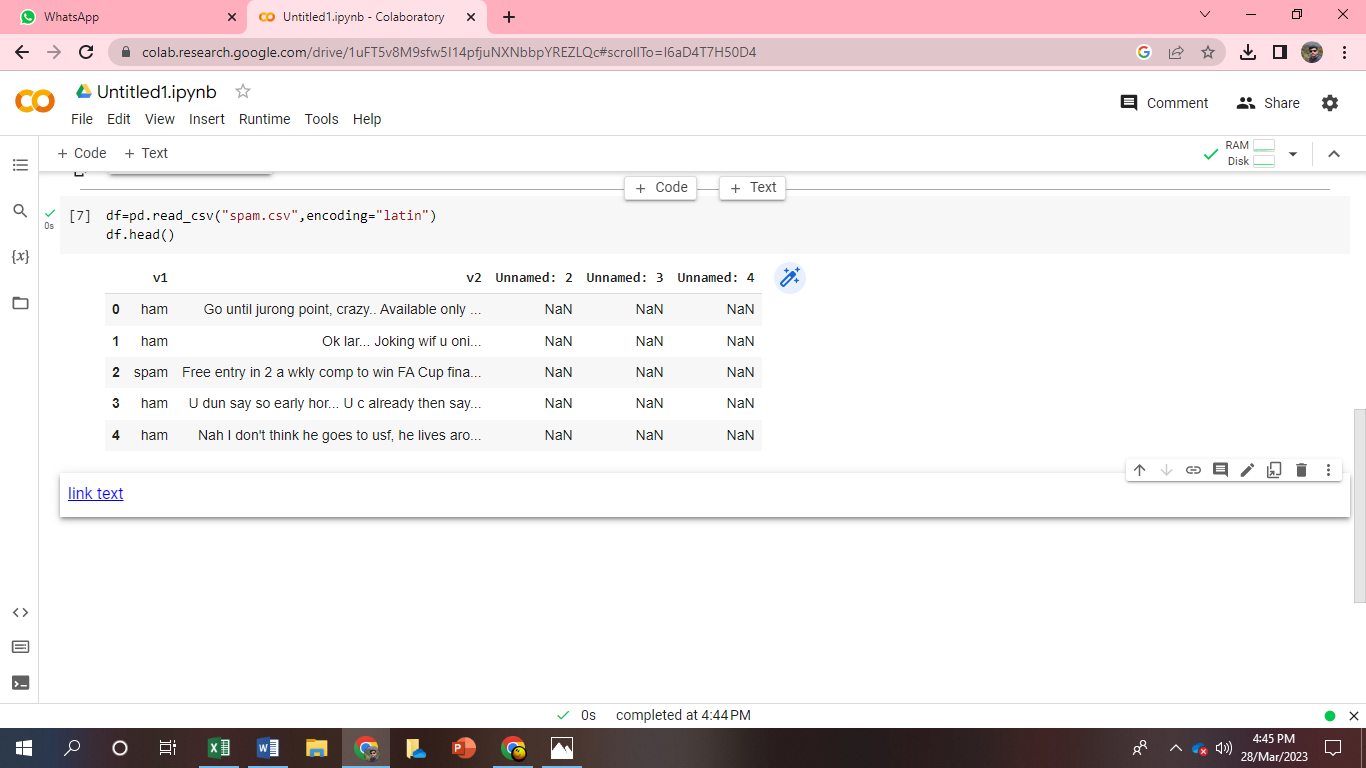
**2.2 Ideation & Brainstorming Map**

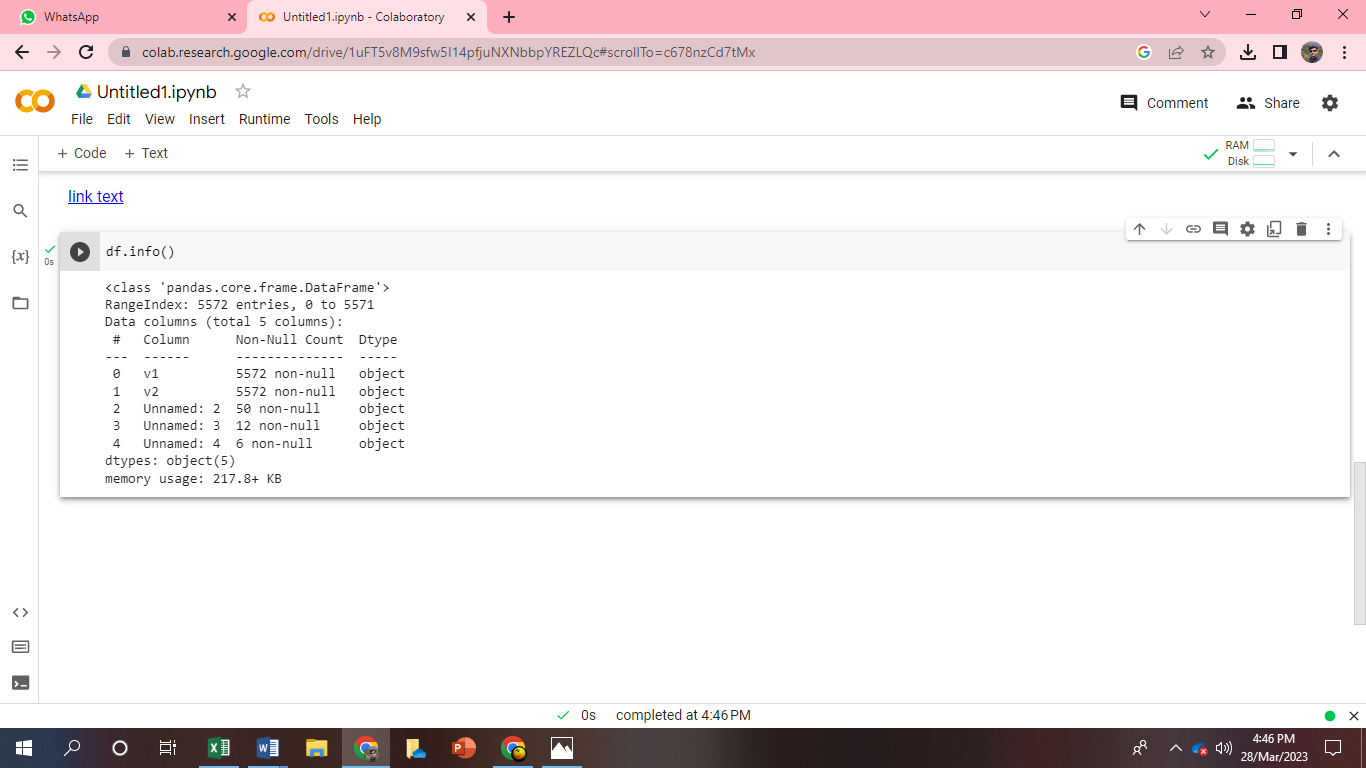


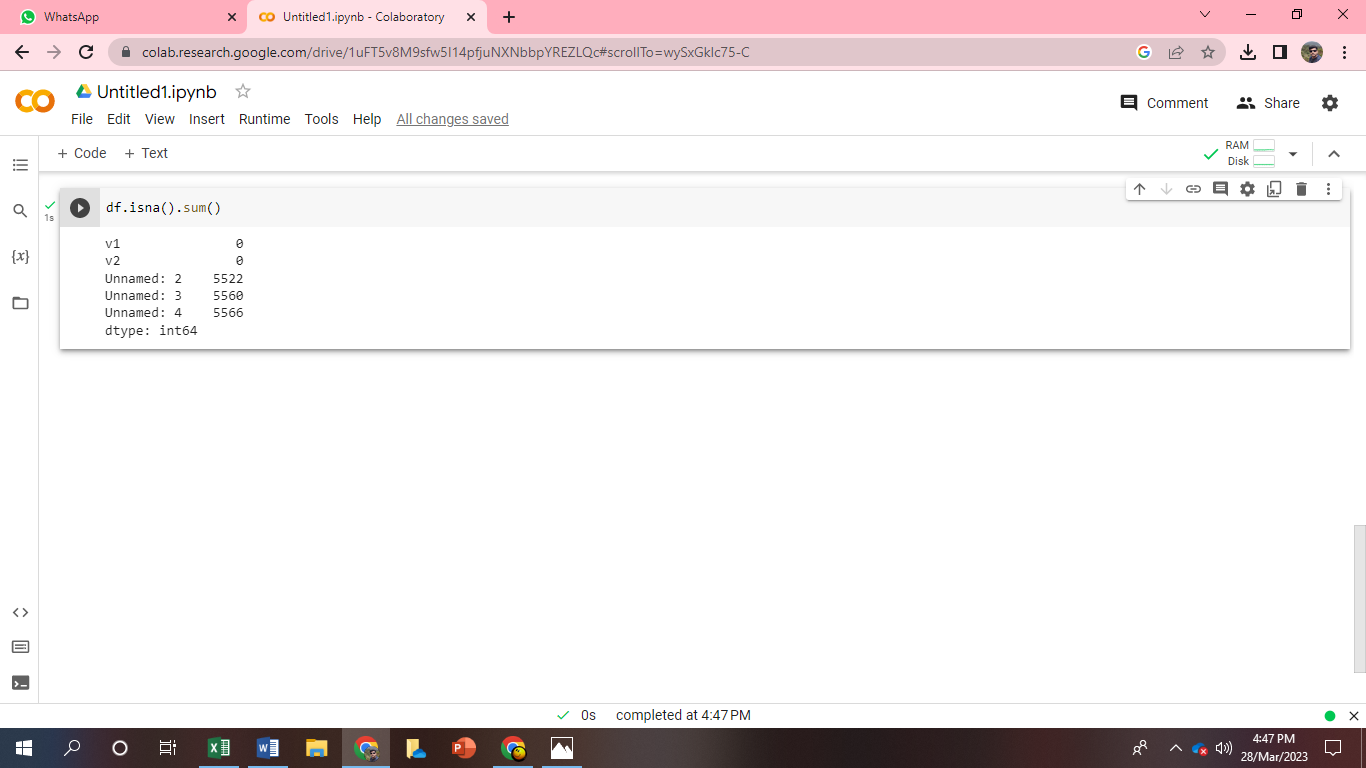


**3. RESULT**









**4. ADVANTAGES AND DISADVANTAGES**

**ADVANTAGES**

Spam is most well-known for spreading viruses and scams to unwitting people across the internet, but it can actually cause plenty of problems for the modern business. This is why effective spam filtering, like Securence spam filtering, is an important part of running a successful business in the 21st century. Here are just a few reasons why spam filtering is important for not only keeping you safe from viruses, but also for helping your company be more effective and successful.

**1. It Streamlines Inboxes**

The average office worker receives roughly 121 emails per day, half of which are estimated to be spam. But even at 60 emails a day, it is easy to lose important communications to the sheer number that are coming in. This is one of the secret benefits of spam filtering that people do not know about: it simply streamlines your inbox. With less garbage coming into your inbox, you can actually go through your emails more effectively and stay in touch with those who matter.

**2. Protect Against Malware**

Malware, viruses, and other forms of malicious attacks are heading to people’s email inboxes every day. Some of these can be easily weeded out by your internet provider’s own spam filters, but spam gets smarter every day. Smarter spam gets into more inboxes, which makes it more likely to be opened and more likely to cause harm. With [spam filtering](http://www.securence.com/spam-filtering-technology/), you can stay on top of the many spam tactics that are being used today so you can ensure that your email inboxes stay free of harmful messages.

**3. Keeps You Compliant**

Many small and medium sized businesses are losing out on important clients today because their cybersecurity is not up to par. Spam filtering is a major part of any cybersecurity plan, and it helps you stay compliant with the wishes and demands of companies and agencies that are concerned about their information. Without proper spam filtering, you could unwittingly put spyware in your emails and break security protocols. The result could be a loss of business, reputation, and ultimately income.

**4. It Saves You Money**

Every day, someone falls prey to a phishing scam, a particular kind of spam-based scheme where someone thinks they are getting a legitimate email and ends up divulging credit card information. Sometimes it is a personal credit card, sometimes it is a company credit card. In both instances, the end result is losing valuable time and money to a scam.

Spam filtering is also incredibly affordable, making it a cheap but extremely effective way to keep yourself safe.

Your inboxes need to be effective tools for communication, not a place where anyone can get into and start hitting you with useless or dangerous emails. That is why spam filtering is such an important aspect of modern businesses. Rather than relying on outdated, free spam filtering services, [choose Securence](http://www.securence.com/) spam filtering.

**DISADVANTAGES**

Most people naturally become nervous as they see they got an email that is unwanted and that is instantly labeled as being spam. If you are a heavy email user there is a pretty good possibility you actually get hundreds of spam emails per week, or even per day, based on activity.

The common approach is to select the unwanted messages and then delete them. However, this is not always something that you want to do. What happens if you mistakenly identified a message as being spam? If this is the case, it is possible you will end up with some serious problems. As an example, if the boss sends you an email and you do not even open it, you might get fired.

**Why Not Use Spam Filters**

The biggest disadvantage of using an email filter is that you may end up with messages being identified as being spam through a mistake of the algorithm that is used. According to [Steven Scott Bayesian](http://people.ischool.berkeley.edu/~hal/Papers/2012/fat.pdf) specialist, even with the very best spam filters on the market you can still end up with messages being improperly labeled.

While missing out on important emails is a nuisance, we need to think about the fact that you can also miss the same emails if you receive a lot of spam. How can you see that message from the boss if there are hundreds of emails sent every single day? You can be highly attentive and still miss out on some emails.

**What Should You Do?**

We cannot dismiss the fact there is a clear disadvantage of using the email spam filters but at the end of the day, it is much better to actually have one installed on the server you use or in the program you use. This means you need to be a little careful and you need to do some things.

For starters, be sure that you use a really good spam filter for your email address. There are many cases in which servers already have some installed but this is not always the case or the filter added is not of a good quality. Ask questions and see what is currently installed. In most situations, the best spam filter you want to look out for is the Bayesian one. Learn the advantages and disadvantages of the email filters before you install them.

The other thing that you can do is to keep using the spam filter. Many just have a filter added to their addresses, like the automatic one that is already a part of Gmail. However, they just do not use them. If you receive an email message that is spam, appropriately label it. Also, from time to time do check the spam tag and see what was automatically added there. Identify the messages that should not be there and “Mark Them As Not Spam”. By doing this you allow the spam filters to adapt and learn. In the future, identifying the spam messages becomes more and more effective. This cannot happen if you just leave the filter to work alone.

**5. APPLICATIONS**

The upsurge in the volume of unwanted emails called spam has created an intense need for the development of more dependable and robust antispam filters. [Machine learning](https://www.sciencedirect.com/topics/computer-science/machine-learning) methods of recent are being used to successfully detect and filter spam emails. We present a [systematic review](https://www.sciencedirect.com/topics/psychology/systematic-literature-review) of some of the popular machine learning based email spam filtering approaches. Our review covers survey of the important concepts, attempts, efficiency, and the research trend in spam filtering. The preliminary discussion in the study background examines the applications of [machine learning techniques](https://www.sciencedirect.com/topics/computer-science/machine-learning-technique) to the email spam filtering process of the leading [internet service providers](https://www.sciencedirect.com/topics/computer-science/internet-service-provider) (ISPs) like Gmail, Yahoo and [Outlook](https://www.sciencedirect.com/topics/nursing-and-health-professions/angiographic-catheter) emails spam filters. Discussion on general email spam filtering process, and the various efforts by different researchers in combating spam through the use machine learning techniques was done. Our review compares the strengths and drawbacks of existing [machine learning approaches](https://www.sciencedirect.com/topics/computer-science/machine-learning-approach) and the open research problems in spam filtering. We recommended deep leaning and deep adversarial learning as the future techniques that can effectively handle the menace of spam emails.

**6. CONCLUSION**

Present a review on supervised machine learning strategies for filtering spam emails. They concluded that **the Naïve Bayes method provides faster results and decent precision over all other methods (except SVM and ID3) from all the techniques discussed**.

Bayesian spam filtering is an incredibly powerful statistical technique—with acceptable computational complexity—for identifying spam messages. Bayesian techniques address many weaknesses of other methodologies:

* The entire message can be examined, not just special parts.
* All words are significant, not just special keywords or addresses.
* Updating is, in practice, infrequent (never more than one or two email messages per week through the training program; often none).
* So far, spam attacks on Bayesian filters have been relatively unsuccessful.
* When combined with other techniques, Bayesian filters can be a very strong component of an institution’s global spam system.

**7. FUTURE SCOPE**

First of all, it’s safe to say that the spam filters that are at work protecting today’s email accounts will continue to improve, until they’re operating at a near 100 per cent success rate. You’ll be able to spend less time sifting through potentially spammy messages, because they simply aren’t going to make it to your inbox in the first place.

Of course, spammers’ tactics are sure to evolve with the times as well, making it difficult to catch every single attempt at spam. However, it’s perhaps even more important that the spam filters of the future become better at identifying those messages that are not considered spam.

As an example, consider that some [7 per cent of emails sent by non-profit organisations today are incorrectly identified as spam](http://www.thenonprofittimes.com/news-articles/efundrasing-appeals-going-spam/). That’s not a terribly high figure, but it’s unfortunate that these well-meaning institutions have to be penalised due to the unscrupulous activities of others.

In the future, spam filters will be more intelligent than ever, correctly distinguishing safe emails from those that need to be removed from the inbox. And that’s going to be possible thanks – in large part – to the ongoing work of companies like MailCleaner.

To be fair, virtual and augmented reality are sure to become a part of virtually everything that we do with our computers, mobile devices and Internet-based activities. All the same, we’re already seeing hints of how our struggle against spam might look in a more virtual environment.

As an example, take Microsoft’s cutting-edge HoloLens – an augmented reality (AR) platform that is just now entering the developer’s market. Microsoft recently released an AR version of Outlook Mail for the headset.

The major innovation of this platform is that it allows you to free up your desktop by moving your emails into an augmented-reality space. You could, for example, move your spam folder to a virtual position on the wall to the side of your desk, so that you can glance over there to see what’s happening in the folder without allowing it to clutter your actual desktop space.

**7. APPENDIX**

**SOURCE CODE**

