SMS SPAM DETECTION

USING IBM WATSON

Kapil Chaturvedi

Electronics and Communication

Vellore Institute of Technology

Vellore, India

kapil.chaturvedi2019@vitstudent.ac.in

Shaswat Pandey

Cyber Security and Digital Forensics CSE

Vellore Institute of Technology

Bhopal,India

shashwat.pandey2019@vitbhopal.ac.in

Aditi Srivastava

Cyber Security and Digital Forensics CSE

Vellore Institute of Technology

Bhopal, India

aditi.srivastava2019@vitbhopal.ac.in

Ananya Saxena

Integrated M.tech specialization in AI and ML

Vellore Institute of Technology

Bhopal, India

ananya.saxena2019@vitbhopal.ac.in

TABLE OF CONTENTS

Table of Contents

1. INTRODUCTION 3
1.1 OVERVIEW 3
1.2 PURPOSE 4
2. LITERATURE SURVEY 5
3. THEORETICAL ANALYSIS 6
1. DATASET 6
2. TEXT PREPROCESSING 6
Countvectorizer *7*
TF-IDF Vectorizer *7*
Hashing Vectorizer *7*
3. MACHINE LEARNING ALGORITHMS 7
1. MultinomialNB *7*
2. SVM *8*
3. Decision Tree *9*
- DATASET 10
- R-squared Value 10
- WEB APP 11

ADVANTAGES: 14
DISADVANTAGE: 14
APPLICATIONS 15
FUTURE SCOPE 16
CONCLUSION 16

1. INTRODUCTION

1.1 OVERVIEW

SMS (Short Message Service) is a standard of mobile protocols which allows users to communicate without connecting to the internet. Due to the cheap cost of SMS services in most telecommunications service providers, and its accessibility and efficiency compared to email services, it is one of the most common communication tools worldwide. However, the attention focused on this service attracts criminals to use it as a means for performing malicious activities and has created trouble for customers and service providers.

An SMS spam message is an unwanted or unsolicited text message which is sent to the user's mobile phone with various content types, such as advertisements, awards, free services and promotions. The main goal of spammers is to steal critical user information such as username, password, and credit card details.

Therefore, in this project, we proposed an Artificial Intelligence method for detection of SMS spam messages. The proposed model contains two main stages: feature extraction and decision making. In the first stage, we have extracted relevant features from the dataset based on the characteristics of spam and legitimate messages to reduce the complexity and improve performance of the model. Then, an averaged neural network model was applied on extracted features to classify messages into either spam or legitimate classes. The method is evaluated in terms of accuracy and F-measure metrics on a realworld SMS dataset with over 5171 messages.

We used various algorithms such as Multinomial Naive Bayes, Support Vector Machine-sigmoid, Support Vector Machine-rbf and Decision tree and found that Multinomial Native Bayes gives the highest accuracy of 96.81 %.

1.2 PURPOSE

A spammer is a person/company which is responsible for unsolicited messages. For their organization benefits or personal benefits, spammers send a vast number of messages to the users. These messages are called spam messages. Although there are various SMS spam filtering techniques available, still there is a need to handle this problem with advanced techniques. Mobile users may get annoyed because of spam messages. Generally, these spam messages are spent by spammers for the promotion of their utilities or business. Sometimes, the users may also undergo financial loss due to these spam messages.

Spammers may impose various strategies in order to steal information, hence, SMS messages being one the most straightforward tactics. One of the most frequent online attacks is phishing which usually happens through email, but the simplicity and extensive usage of mobile phones have made phishers consider SMS messages as a suitable method. In phishing attacks, the phisher sends a malicious URL using SMS messages and invites users to visit that URL address in order to steal sensitive and personal information from the user's mobile phone. Moreover, SMS phishing has no severe limitation for spammers, and they can easily buy various phone numbers within any area or country code to send malicious SMS messages. This makes it challenging to recognize and distinguish attackers based on their mobile number.

Accordingly, a reliable and accurate method to filter spam messages is essential, although different security techniques are available to block these messages. So we introduced an AI method for classification of SMS spam messages in order to alleviate the challenges and detect spam messages effectively with a high detection rate.

2. LITERATURE SURVEY

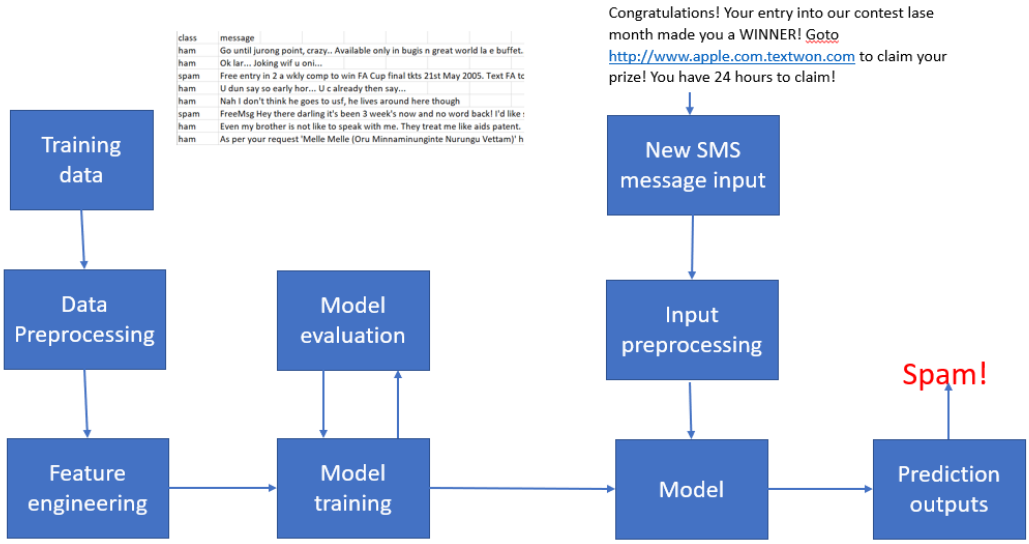
The goal of this survey is to undertake a thorough literature evaluation on approaches for detecting and classifying spam content in SMS. Spam text has been detected and regulated in a variety of methods. Our efforts are primarily motivated by a desire to gain a better understanding of various spam SMS detection methods.

Many cellular operators have recently devised sophisticated mechanisms, like Open Mobile Alliance (OMA), to protect mobile users from SMS spam. While these administrative measures cut off a massive quantity of spam directed towards consumers, the problem of personalized spam filtering on mobile devices remains unsolved. The existing spam filtering techniques for mobile phones are based on the content of SMS.

Applying ML techniques for spam detection is not a new era. Previously, various researchers applied ML techniques for classification of SMS spam. Nilam Nur Amir Sjarif[1] et.al applied the TF-IDF technique in combination with a random forest classifier and achieved an accuracy of 97.5%.TF-IDF is a method used to quantify the words in a document by using two measures Term Frequency and Inverse Document Frequency.A.Lakshmana Rao[2] et.al applied four machine learning classifiers Decision Trees, Naive Bayes, Logistic Regression, Random Forest for email spam filtering, and achieved an accuracy of 97% with random forest classifier. Pavas Navaney[3] et.al proposed various machine learning algorithms and achieved an accuracy of 97.4% with support vector machines.

Most of these techniques are straight forward adaptations of email spam detection schemes and usually incorporate features–specific words, character bi-grams, and tri-grams–for the classification of spam messages. Some other works have proposed schemes based on machine learning algorithms for the classification of SMS spam. We now present a SMS spam detection framework that works at the access layer of a mobile phone to detect spam messages.

3. THEORETICAL ANALYSIS



1. DATASET

We took a dataset that contains 5172 rows with two columns. The first column specifies whether the message is “spam” or “ham”. Here “spam” means unsolicited message and “ham” means normal message. The second column contains the actual message.

2. TEXT PREPROCESSING

Text data can be represented in vectorized format. Text preprocessing can be done by various NLP(Natural Language Processing) techniques. Machine Learning algorithms work with numbers only. So, there is a need to encode text data into the numeric format. Tokenization is a process of dividing text data into different parts. In-Text preprocessing stop words are removed. stop words are the words that are not useful for analyzing the text data. For example, the words like is, was, that, are the stop words. After removing, stemming can also be applied. Stemming is a process of reducing the word to its stem. For example, the word “playing” can be changed as “play”. After that word embeddings can be done, where the words are changed as vectors of real values.

We applied three different word embedding techniques:

• Countvectorizer

• TF-IDF Vectorizer

• Hashing Vectorizer.

Countvectorizer

In this, first, all the preprocessing is done like removing special symbols, converting to lowercase, etc. Later, it identifies the unique words in the whole text and creates an array of zeros for every sentence. Next, it adds word count to each sentence. The resultant vector is the vector representation of the given text.

TF-IDF Vectorizer

This model uses two measures Term Frequency and Inverse Document Frequency.TF is the number of times the sentence appears.IDF is inverse document frequencies.

Hashing Vectorizer

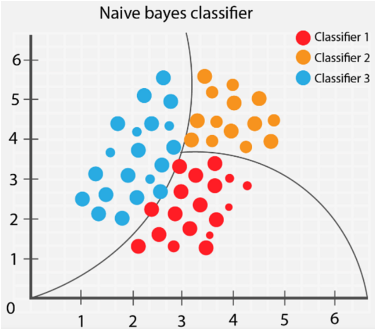
Hashing Vectorizer uses a hashing algorithm for converting text to vector.The algorithm can be applied to all sentences in the document.

3. MACHINE LEARNING ALGORITHMS

1. MultinomialNB

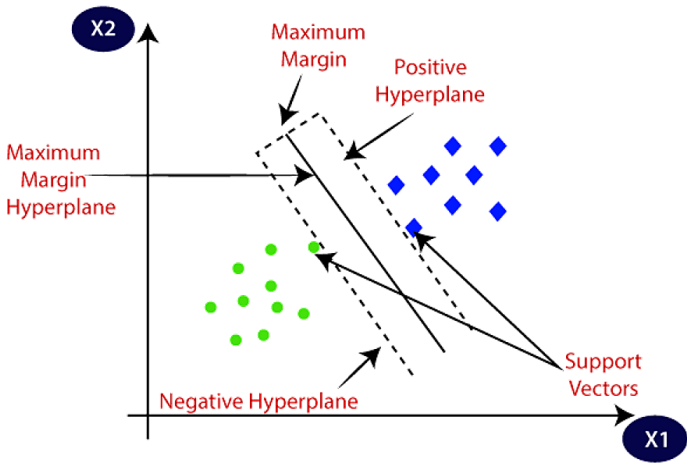
Multinomial Naive Bayes algorithm is a probabilistic learning method that is mostly used in Natural Language Processing (NLP). The algorithm is based on the Bayes theorem and predicts the tag of a text such as a piece of email or newspaper article. It calculates the probability of each tag for a given sample and then gives the tag with the highest probability as output.

Naive Bayes classifier is a collection of many algorithms where all the algorithms share one common principle, and that is each feature being classified is not related to any other feature. The presence or absence of a feature does not affect the presence or absence of the other feature.



2. SVM

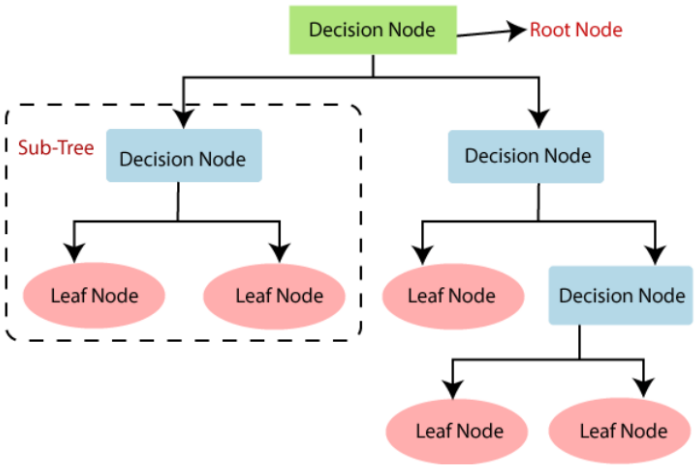
Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. The objective of the SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.



3. Decision Tree

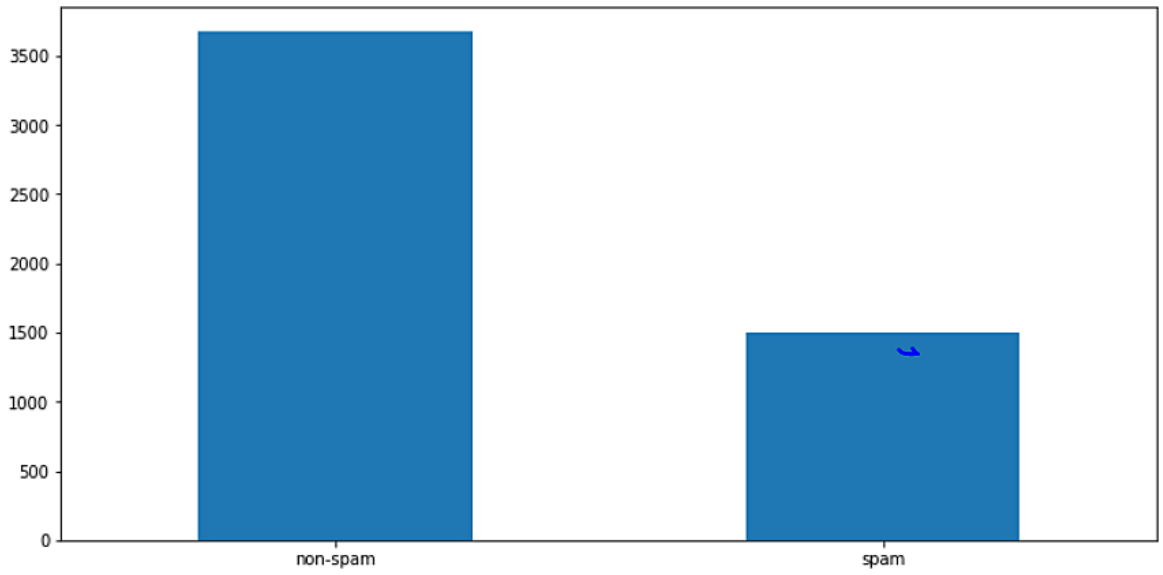
Decision tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

A tree can be *"learned"* by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called *recursive partitioning*. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. The construction of a decision tree classifier does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery.



RESULT

- **DATASET**

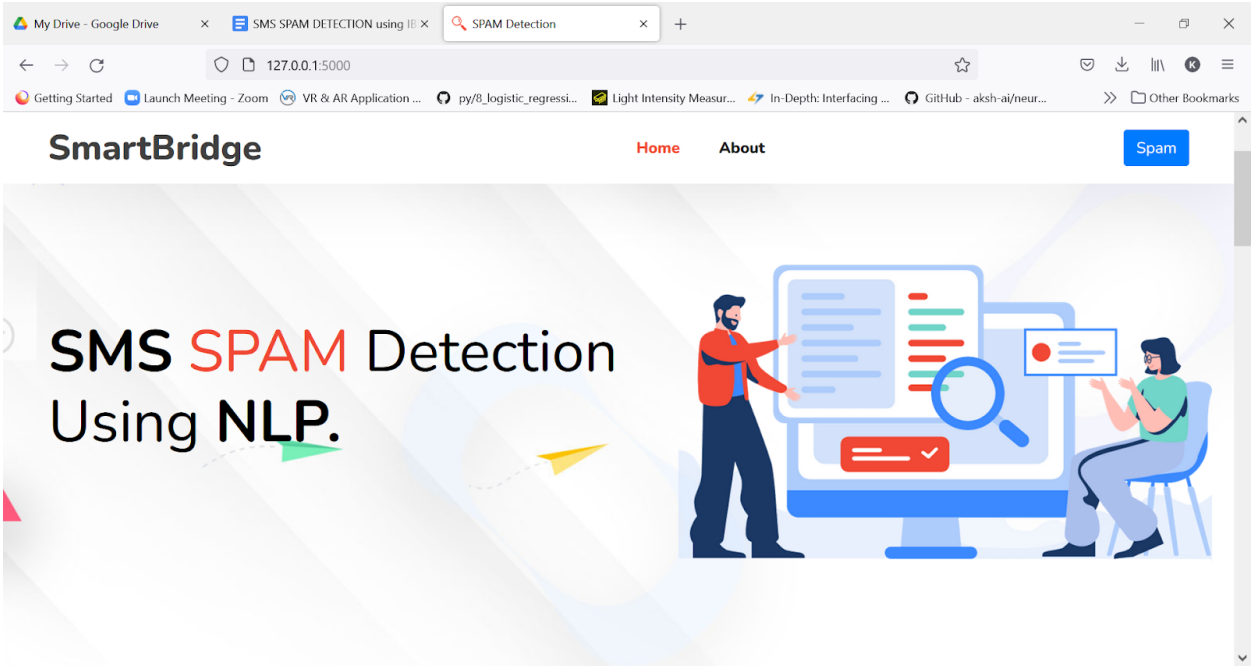


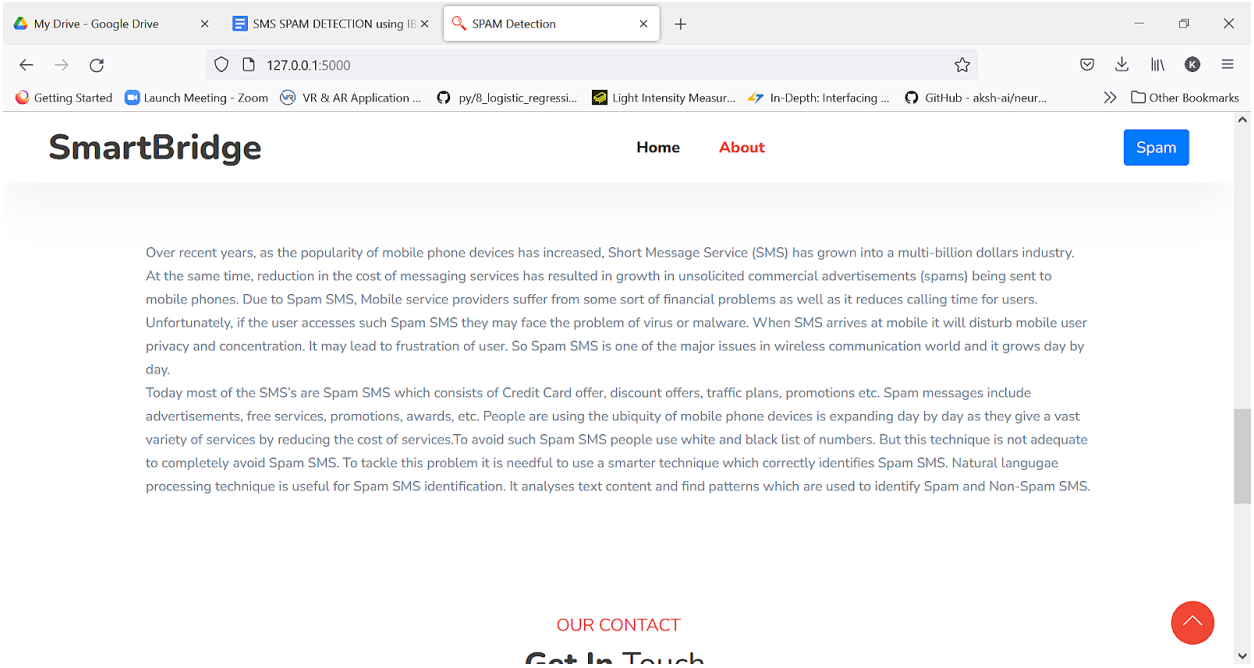
- **R-squared Value**

	Model	Test Score
0	MultinomialNB	0.968116
2	SVM-sigmoid	0.965217
1	SVM-rbf	0.962319
3	Decision Tree	0.945894

MultinomialNB gives the maximum accuracy while predicting “spam” which is an unsolicited message or “ham” which is the normal message with a R-squared value of 0.968116.

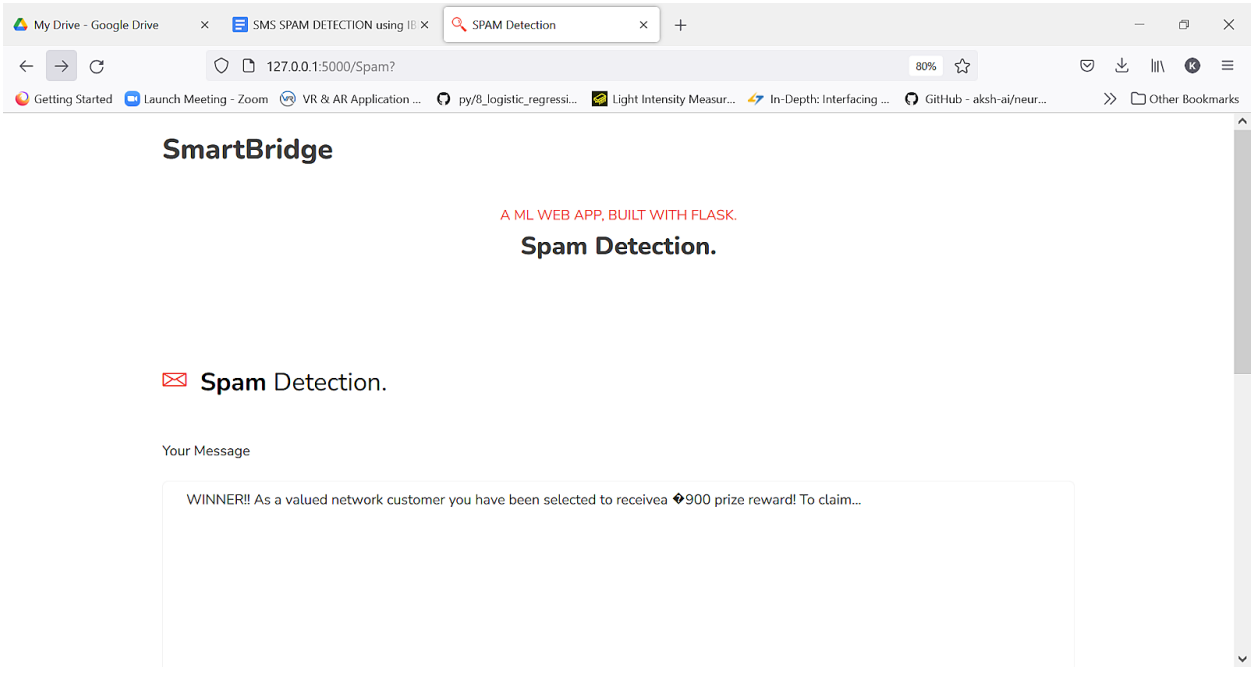
- **WEB APP**

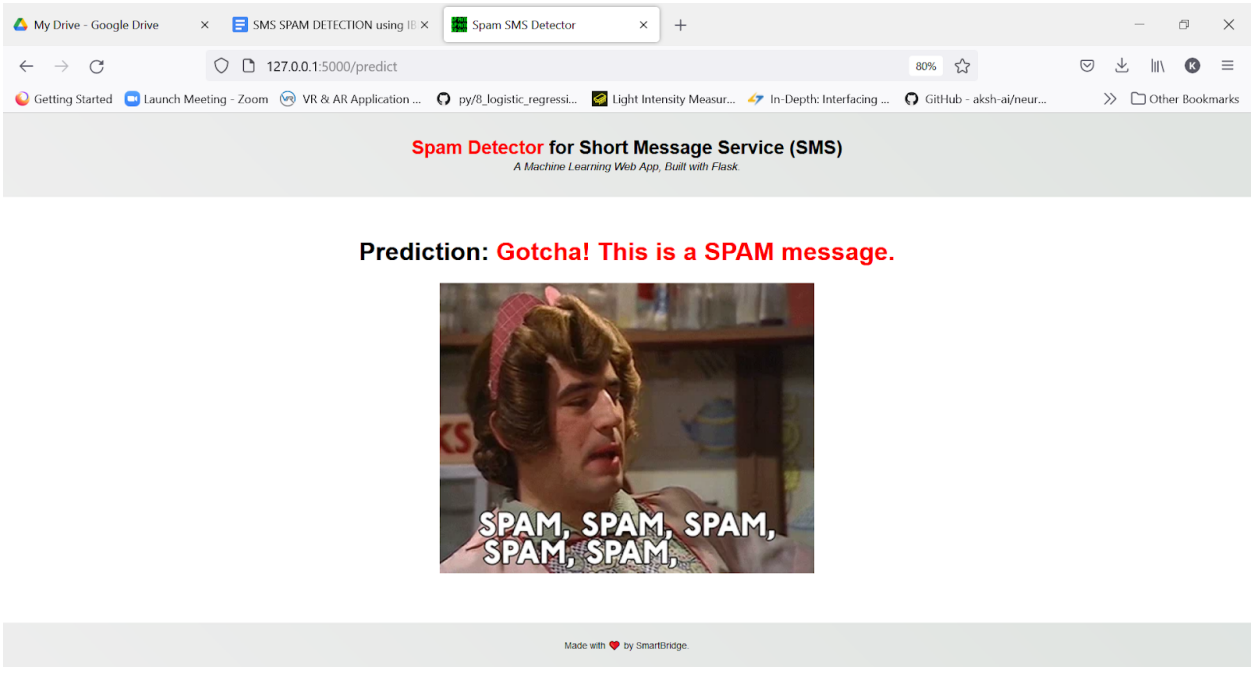

My Drive - Google Drive
SMS SPAM DETECTION using
SPAM Detection
127.0.0.1:5000
Getting Started
Launch Meeting - Zoom
VR & AR Application ...
py/8_logistic_regressi...
Light Intensity Measur...
In-Depth: Interfacing ...
GitHub - aksh-ai/neur...
Other Bookmarks
SmartBridge
Home
About
Spam
SMS SPAM Detection
Using NLP.


My Drive - Google Drive
SMS SPAM DETECTION using
SPAM Detection
127.0.0.1:5000
Getting Started
Launch Meeting - Zoom
VR & AR Application ...
py/8_logistic_regressi...
Light Intensity Measur...
In-Depth: Interfacing ...
GitHub - aksh-ai/neur...
Other Bookmarks
SmartBridge
Home
About
Spam
Over recent years, as the popularity of mobile phone devices has increased, Short Message Service (SMS) has grown into a multi-billion dollars 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 of user. So Spam SMS is one of the major issues in wireless communication world and it grows day by day.
Today most of the SMS's are Spam SMS which consists of Credit Card offer, discount offers, traffic plans, promotions etc. Spam messages include advertisements, free services, promotions, awards, etc. People are using the ubiquity of mobile phone devices is expanding day by day as they give a vast variety of services by reducing the cost of services.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 langugae processing technique is useful for Spam SMS identification. It analyses text content and find patterns which are used to identify Spam and Non-Spam SMS.
OUR CONTACT

- **SPAM Message**

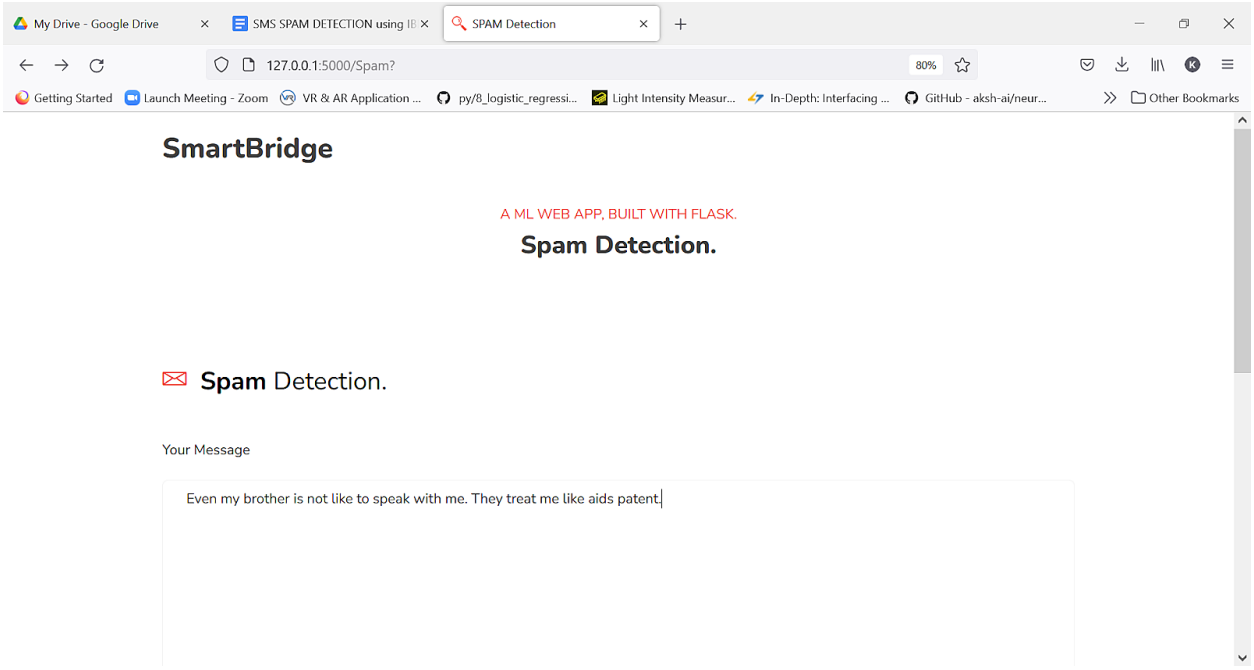
  (WINNER!! As a valued network customer you have been selected to receive a ₹900 prize reward! To claim…)

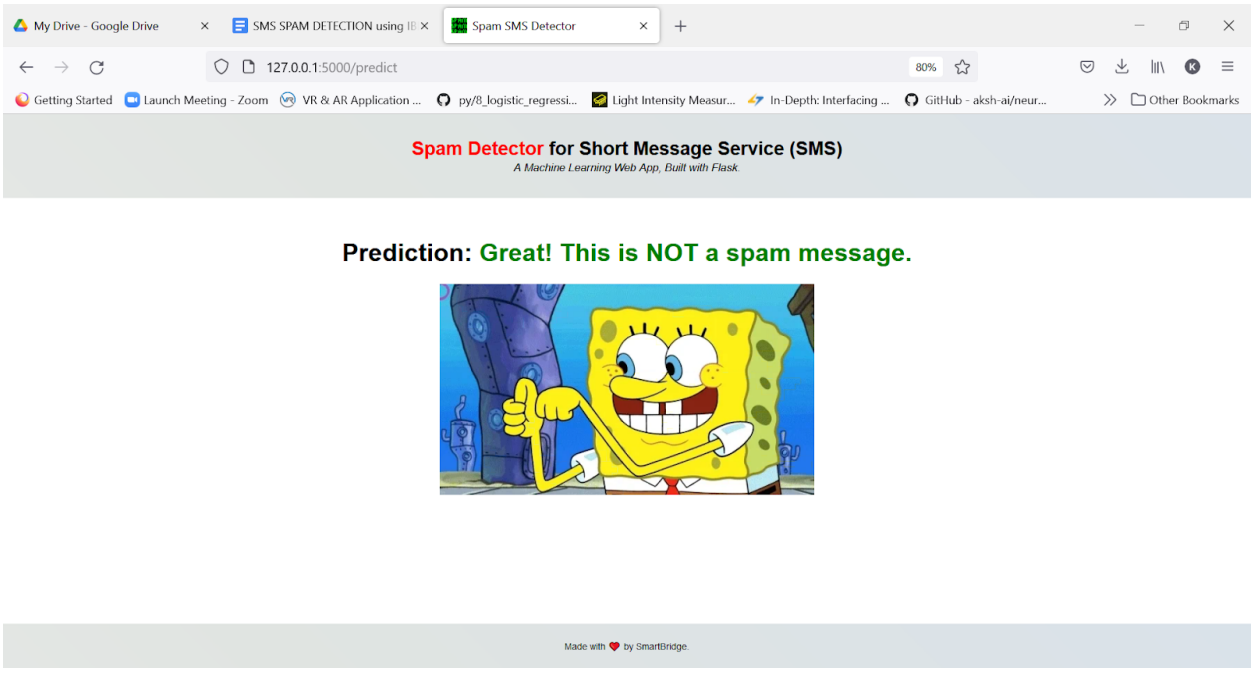




- **HAM Message**

  (Even my brother is not like to speak with me. They treat me like aids patent……)





ADVANTAGES:

1. **It Streamlines Inboxes -** On an average an individual sends/receives around 70 texts per day out of which 60% are spam messages. 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 our SMS 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 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, 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. **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 message 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.

DISADVANTAGE:

- The biggest disadvantage of using an SMS Spam filter is that we may end up with messages being identified as being spam through a mistake of the algorithm that is used. According to Steven Scott Bayesian specialist, even with the very best spam filters on the market wecan still end up with messages being improperly labeled.

- A subtle issue with Naive-Bayes Classifier is that if you have no occurrences of a class label and a certain attribute value together then the frequency-based probability estimation will be zero.
- A big data set is required for making reliable predictions of the probability of each class. We can use this with small data sets but the precision will be altered.

APPLICATIONS

- A spam filtering solution cannot be 100 percent effective. However, a business email/SMS system without spam filtering is highly vulnerable, if not unusable. It is important to stop as much spam as you can, to protect your network from the many possible risks: viruses, phishing attacks, compromised web links and other malicious content.

- Spam filters also protect our servers from being overloaded with non-essential emails/SMS, and the worst problem of being infected with spam software that may turn them into spam servers themselves.

- By preventing spam email and SMS from reaching our employees' mailboxes, spam filters give an additional layer of protection to our users, your network, and our business.

- In Google Voice, Google has a global spam filtering system that automatically redirects calls, texts and voicemails from any of the numbers listed in Google's global spam database. Users must enable this feature in the Calls tab of Google Voice settings.

- The Linux cPanel interface enables the configuration of spam filter settings for user accounts in web hosting environments. Users can also use cPanel to configure and manage their blocklist and allowlist settings.

- Other platforms, like social media giant Instagram use spam filters to help maintain order. Instagram users can also use the spam filter to block harassment and unwanted direct message requests.

FUTURE SCOPE

This project was a study-case about developing an SMS spam classification model based on various algorithms. For future works, we aim to enhance the performance of the model by collecting more data from various sources to develop the model. We expect to develop a model that can be used to help people in the real-world.

Here we proposed an AI model for prediction of spam based on the content data like only message content. In future work we can elaborate this topic to prediction by using content and context data like Host address of the SMS, sender, number of times received, URL‟s in the messages etc.

CONCLUSION

Detection of spam is important for securing messages. The accurate detection of spam is a big issue, and many detection methods have been proposed. To solve this issue, we have proposed a method for spam detection using machine learning predictive models. The method is applied for the purpose of detection of spam. The experimental results obtained show that the proposed method has a high capability to detect spam. The proposed method achieved 96.81% accuracy. Thus, the results suggest that the proposed method is more reliable for accurate and on-time detection of spam, and it will secure the communication systems of messages.

The accuracy of the proposed model can be improved further on evaluation using bigger datasets and finding a standard sizable dataset which was one of the limitations in this research.

REFERENCES

1. Nilam Nur Amir Sjarif, N F Mohd Azmi, Suriayati Chuprat, “SMS Spam Message Detection using Term Frequency-Inverse Document Frequency and Random Forest Algorithm,” in The Fifth Information Systems International Conference 2019, Procedia Computer Science 161 (2019) 509-515,ScienceDirect.

2. A.Lakshmana Rao,K.Chandra Sekhar, Y.Swathi, “An Efficient Spam Classification System using Ensemble Machine Learning Algorithm,” in Journal of Applied Science and Computations, Volume 5, Issue 9, September/2018.

3. Pavas Navaney, Gaurav Dubey, Ajay Rana, “SMS Spam Filtering using Supervised Machine Learning Algorithms.,” in 8th International Conference on Cloud Computing, Data Science & Engineering, 978-1- 5386-1719-9/18/ 2018 IEEE

4. Shirani-Mehr, H. (2013). SMS spam detection using machine learning approach. *unpublished) http://cs229. stanford. edu/proj2013/Shir aniMeh*

5. M. Bassiouni, M. Ali, and E. A. El-Dahshan, “Ham and spam E-mails classification using machine learning techniques,” Journal of Applied Security Research, vol. 13, no. 3, pp 315–331, 2018.

6. Navaney, P., Dubey, G., &Rana, A. (2018). "SMS Spam Filtering Using Supervised Machine Learning Algorithms." 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence).

7. Yang, Y., J.O. Pedersen. 1997. A comparative study on feature selection in text categorization. En Proceedings of the 14th International Conference on Machine Learning.