

INTERDISCIPLINARY PROJECT REPORT
at
Sathyabama Institute of Science and Technology
(Deemed to be University)

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering Degree in Computer Science and Engineering

By

NAME:S.HARSHA VARDHAN REDDY
(Reg.No.40733024)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING
SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119, TAMILNADU

APRIL 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

(Established under Section 3 of UGC Act, 1956)

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI- 600119

www.sathyabamauniversity.ac.in



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **V. MEGHA SYAM (41111301)** who carried out the project entitled "**Smart car based on IOT**" under my supervision from FEB 2023 to APRIL 2023.

Internal Guide

Name: Ms.K.ABHINAYA MAM

Head of the Department

Dr.L.LAKSHMANAN M.E., Ph.D.

Submitted for Viva voce Examination held on 17-04-2023

Internal Examiner

External Examiner

DECLARATION

IS.Harsha Vardhan Reddy hereby declare that the Project Report entitled Cold SMS Filtering

Done by me under the guidance of Dr./Prof./Mr./Ms. _____ (Internal) and _____ (External)

at SMART BRIDGE (Company name and address) is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE: 17-04-2023

PLACE: Chennai



SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D., Dean**, School of Computing , **Dr.S.Vigneshwari M.E., Ph.D., and Dr.L.Lakshmanan M.E., Ph.D.,** Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide -----
-----for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my projectwork.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

TRAINING CERTIFICATE

ABSTRACT

This paper presents detection of Spam and ham messages using various supervised machine learning algorithms like naïve Bayes Algorithm, support vector machines algorithm, and the maximum entropy algorithm and compares their performance in filtering the Ham and Spam messages. As people indulge more in Web-based activities, and with rising sharing of private – data by companies, SMS spam is very common. SMS spam filter inherits much functionality from E-mail Spam Filtering. Comparing the performance of various supervised learning algorithms we find the support vector machine algorithm gives us the most accurate result.

Cold SMS filtering refers to the process of automatically identifying and separating unsolicited text messages from legitimate ones. This task is becoming increasingly important as the number of unwanted SMS messages sent by spammers and scammers continues to rise. In this process, various filtering techniques such as rule-based filtering, machine learning-based filtering, and content-based filtering are utilized to effectively detect and block unwanted SMS messages. The aim of this process is to ensure that only genuine and relevant messages are received by the users, thereby protecting them from fraudulent activities and improving their overall experience.

Cold SMS filtering is a crucial aspect of mobile messaging security that helps protect users from spam and other types of unwanted messages. Spam text messages can be annoying and intrusive, and they can also pose a security risk by attempting to trick users into revealing personal information or installing malicious software.

There are several different approaches to filtering cold SMS messages. One method is to use rule-based filters, which work by applying a set of predefined rules to incoming messages to determine whether they are spam or legitimate. These rules may be based on factors such as the sender's phone number, message content, or time of day.

Another approach is to use machine learning-based filtering, which involves training algorithms to recognize patterns and characteristics of spam messages based on large amounts of data. These algorithms can then be used to automatically classify incoming messages as either spam or legitimate.

Content-based filtering is another method used for cold SMS filtering. This approach involves analyzing the actual content of messages to identify spam or malicious content. For example, a content-based filter may look for specific keywords or phrases commonly associated with spam, such as "free" or "limited time offer."

Chapter No	TITLE		Page No.
	ABSTRACT		vi
	LIST OF FIGURES		viii
1	INTRODUCTION		
	1.1	Architecture	
	1.2	Flowchart	
	1.3	What is cold texting	
	1.4	Is cold texting is illegal?	
2	AIM AND SCOPE		
	2.1	AIM	15

	2.2	SCOPE How does the message filtering work	15 15-16
3		METHODOLOGY AND ALGORITHMS USED	
	3.1	SYSTEM REQUIREMENTS	
	3.2	PACKAGES USED	

	3.3	ALGORITHM	18
	3.4		
	3.5		
	3.6		
4	RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS		
	4.1	Results	
	4.2	Performance analysis	
5	SUMMARY AND CONCLUSION		
	5.1	Summary	
	5.2	Conclusion	
	REFERENCES		
	APPENDIX		
	A. SOURCE CODE		
	B. SCREENSHOTS		

LIST OF FIGURES

FIGURE No	FIGURE NAME	PAGE No
1.		
2.		
3.		
4.		
5.		

6.		
7.		
8.		

LIST OF ABBREVIATIONS

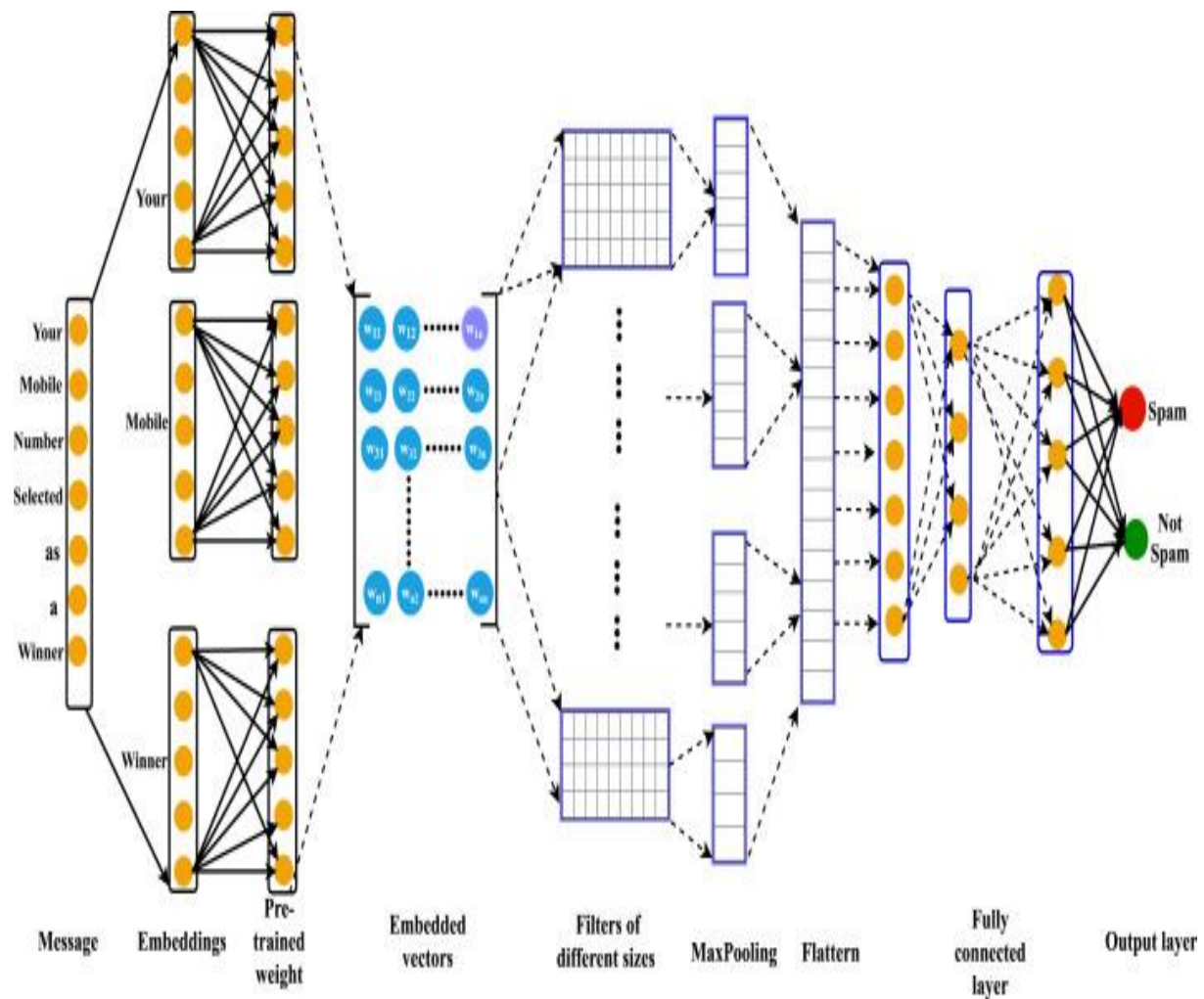
ABBREVIATION	EXPANSION
ML	Machine learning
HTML	Hyper text markup language

NLP	Natural Language processing
PY with ML	Python with machine learning

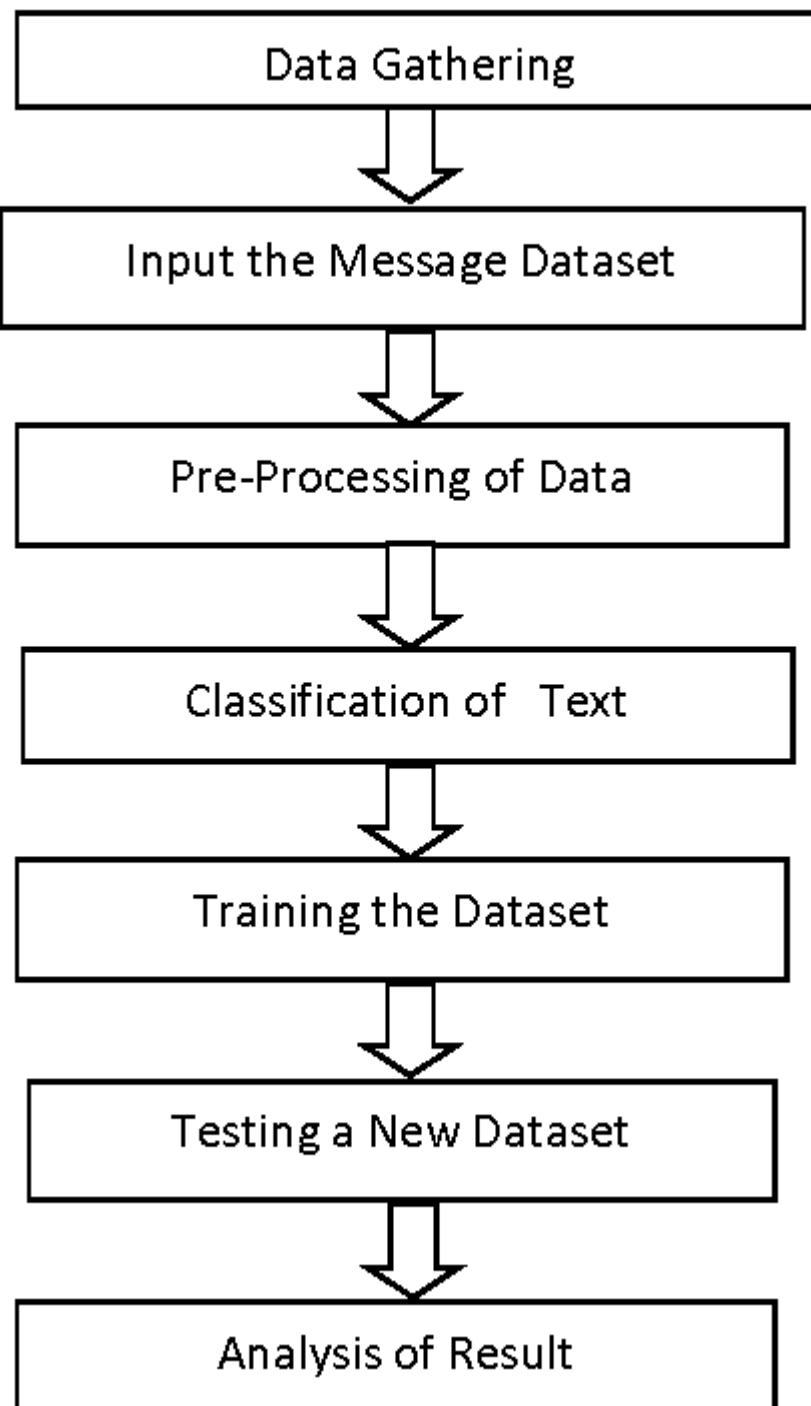
INTRODUCTION

- ▶ In the developing period of the Internet, individuals are involving increasingly in free online services. Individuals tend to share their data on different sites, though that data is imparted to different organizations that spam individuals to offer their services. SMS Spamming [2] [10] is extremely disappointing for the clients: numerous critical and valuable messages can get lost because of spam messages, Spam messages are additionally used to trap individuals, or bait them into purchasing services. As overall utilization of cell phones has grown, another road for e-junk mail has been opened for notorious advertisers. These publicists use instant messages (SMS) to target probable purchasers with undesirable publicizing known as SMS spam. This sort of spam is especially bothersome since, not at all like email spam, numerous PDA clients pay an expense for each SMS got.
- ▶ Building up a classification algorithm [1] [11] that channels SMS spam would give a helpful apparatus for mobile phone suppliers. Since naïve Bayes has been utilized effectively for email spam detection [9], it appears to be expected that it could likewise be used to build SMS spam classifier [7]. With respect to email spam [6][8], SMS spam represents extra difficulties for automated channels. SMS texts are regularly restricted to 160 characters, lessening the measure of content that can be utilized to distinguish whether a message is a ham or spam. People have also regularly started using shorthand notations and slang which further makes it difficult to distinguish between ham and spam. We will test how well a simple naïve Bayes classifier [4] manages these difficulties.
- ▶ We additionally fabricate models to group messages utilizing the SVM algorithm and the maximum entropy algorithm [3], and it is discovered that SVM gives us the most precise outcomes, with exactness up to 98 %, took after by Naïve bayes algorithm, followed by maximum entropy algorithm.
- ▶ Spam messages can be classified as redundant messages sent to large number of people at once. The rise of spam messages are based on the following factors:
 - 1) The accessibility to cheap bulk SMS-plans
 - 2) dependability(since the message comes to the cell phone client)
 - 3) low possibility of accepting reactions from some unaware recipients and
 - 4) the message can be customized
 - 5) Free services

Architecture of Cold SMS filtering:-



FLOW CHART:-



What is Cold texting?

Cold texting is reaching out via SMS to a contact without their prior consent or permission. It's similar to cold calling and cold emailing.

Some examples that would lead to cold texting include:

- Purchasing a contact list from a third party
- Finding customers' phone information online
- Obtaining someone's phone number in a way they don't know about

Is cold texting illegal?

Yes, cold texting is illegal. The [Telephone Consumer Protection Act](#) requires businesses to get an opt-in from customers to text them. You may face hefty fines if you don't obtain [SMS opt-ins](#).

Stop cold texting and do this instead

Instead of sending a cold text message, organically grow your [SMS subscriber list with these strategies](#).

- **Make your phone number public.** Encourage customers to contact your business first by publicizing your texting number. When customers text in first, they're providing [implied consent](#).
- **Use opt-in forms.** Generate more opt-ins by using online or physical forms. For example, you could create a pop-up form for your website or ask customers to fill out a form in person.
- **Work opt-ins into a current sign-up flow.** Anywhere customers insert their phone numbers, include a check box they can click that explicitly states they give your business permission to reach out via SMS.
- **Ask for consent with keywords.** Use a text-to-join campaign to encourage customers to sign up for marketing texts. These campaigns ask customers to text a keyword like "JOIN" to your number to subscribe. Send a follow-up text to confirm the opt-in choice.
- **Convert social media followers or email subscribers.** Provide a way for customers who follow you on social media or subscribe to your email campaigns to opt into text messaging.
- **Allow SMS opt-outs.** As the law requires, businesses must provide a way for contacts to opt out. Providing a way to opt out may encourage more customers to subscribe since the choice to receive texts is entirely theirs.

- **Stay compliant.** The big texting law to follow is the TCPA. Ensure you're staying compliant with this law by following [this eight-step compliance checklist](#). Compliant texting will help you maintain your SMS subscriber base. And avoid fines and penalties.

How to ensure your email and SMS are delivered

Increase the chances of having your emails and texts safely delivered by following these tips.

Follow the CAN-SPAM rules

The [CAN-SPAM act](#) regulates how emails and messages are sent and to whom. It offers recipients the right to opt-out of business communications with a specific company and specifies what penalties apply if the company doesn't respect the recipient's request.

If you want your texts and emails to pass the CAN-SPAM inspection, you need to:

- Be upfront about what you're selling or asking your clients. Don't lead them on.
- Clearly state which company you represent.
- Write a clear subject line.
- Always offer recipients the option to unsubscribe from your written communications.
- End every communication with a [professional signature](#).

Avoid cold SMS/email marketing

>Contacting people who have no prior relationship with your brand is intrusive and can seriously **damage your company's reputation**.

>Here are a couple of alternatives to cold texting or emailing that will help you establish long-lasting business relationships with customers who truly value what you have to offer:

>Boost your networking efforts and attend more industry events to increase brand visibility.

>Produce content that turns you into a figure of authority in a specific field or niche – start with a blog and expand from there.

>Stand out from the competition with offers they can't resist (create personalized marketing campaigns, use social media to boost marketing efforts).

CHAPTER-2

Aim:-

The aim of the cold SMS filtering is to detect the SPAM and HAM messages .

This guide discusses SMS filtering that can affect delivery of your Twilio Programmable SMS messages. This is done to enforce Twilio's [Messaging Policy](#) or [Acceptable Use Policy](#), and/or to comply with regulations and wireless carriers' messaging policies.

Filtering of your outgoing Twilio Programmable SMS messages can vary somewhat between carriers and countries. We have compiled the following information to help you understand why a message might have been filtered.

Scope:-

>The scope of the project is to understand the basic concept of areas of the text mining. It also involves the text classification Model to perform predictive analysis. In modern day the experiment is the key. The study involves the modern analytics tools. This tools help understand the working knowledge of the different technique which are used in analytics. The modern analytics applications can which can be applied in the real time scenario. The scope of every study is based on data collection methods. Company provided the data for the development of a classification model for SPAM/HAM SMS classification using caret and Naive Bayes.

>This model will help the company to apply for model for predicting different SPAM/HAM SMS in text analytics market.

How Does Message Filtering Work?

This guide discusses SMS filtering that can affect delivery of your Twilio Programmable SMS messages. This is done to enforce Twilio's [Messaging Policy](#) or [Acceptable Use Policy](#), and/or to comply with regulations and wireless carriers' messaging policies.

Filtering of your outgoing Twilio Programmable SMS messages can vary some what between carriers and countries. We have compiled the following information to help you understand why a message might have been filtered.

Message filtering can range from a simple static list of prohibited terms, to advanced machine learning systems that constantly adapt based on the messages passing through them.

Twilio uses filtering to target messaging that is in violation of our [Messaging Policy](#) and [Acceptable Use Policy \(AUP\)](#). Twilio's goal is to protect mobile users from fraud, spam, and other forms of unwanted messaging, as well as to comply with rules and regulations imposed by carriers and regulators.

In addition, carriers filter messages to enforce their rules, as well as to prevent spam and other forms of abuse.

CHAPTER 3

Algorithm:-

Step 1:- importing the modules

Step 2:-Reading the data set

Step 3:-EDA on data set

Step 4:- understanding the data type and Summary features

Step 5:-Take care of missing data

Step 6:- Data Visualization

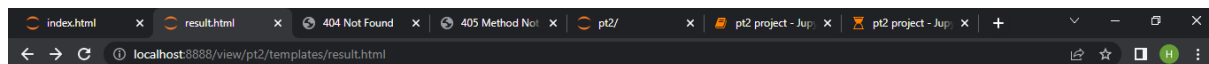
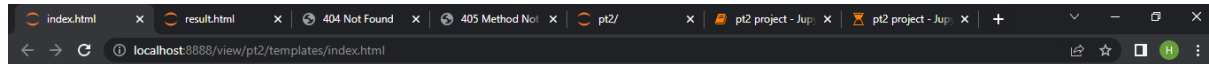
Step 7:- Cleaning the text

CHAPTER 4

Results and discussion

Results:-

The result of this project is to detect the SPAM/HAM SMS messages. This will be going to check the SPAM/HAM messages.



Discussion:-

The above diagram shows the message or SMS which we got is SPAM/HAM or not

Source Code:-

```
from sklearn.naive_bayes import MultinomialNB

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import nltk
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.naive_bayes import MultinomialNB
import re
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.svm import SVC
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.tree import DecisionTreeClassifier
import pickle

import sklearn

df = pd.read_csv("spam_ham_dataset.csv")

df.head()

df.tail()

df=df.drop("Unnamed: 0",axis=1)

df.shape

df.ndim

df.size

df.isna().sum()

df.info()
```

```
df.describe()
```

```
df.isnull().any()
```

```
df.isnull().sum()
```

```
df["label_num"].value_counts().plot(kind="bar",figsize=(12,6))  
plt.xticks(np.arange(2),('Non spam', 'spam'),rotation=0);
```

```
import re  
corpus = []  
length = len(df)
```

```
for i in range(0,length):  
    text = re.sub("[^a-zA-Z0-9]", " ",df["text"][i])  
    text=text.lower()  
    text=text.split()  
    pe = PorterStemmer()  
    stopword = stopwords.words("english")  
    text = [pe.stem(word) for word in text if not word in set(stopword)]  
    text = " ".join(text)  
    corpus.append(text)
```

```
corpus
```

```
from sklearn.feature_extraction.text import CountVectorizer  
cv = CountVectorizer(max_features=35000)  
X = cv.fit_transform(corpus).toarray()
```