Email Spam Detection

Ankita Wagavekar(21CO132)

Student at AISSMS COE.

Abstract

Nowadays, a big part of people rely on available email or messages sent by the stranger. The possibility that anybody can leave an email or a message provides a golden opportunity for spammers to write spam message about our different interests .Spam fills inbox with number of ridiculous emails . Degrades our internet speed to a great extent .Steals useful information like our details on our contact list. Identifying these spammers and also the spam content can be a hot topic of research and laborious tasks. Email spam is an operation to send messages in bulk by mail .Since the expense of the spam is borne mostly by the recipient ,it is effectively postage due advertising. Spam email is a kind of commercial advertising which is economically viable because email could be a very cost effective medium for sender .With this proposed model the specified message can be stated as spam or not using Bayes' theorem and Naive Bayes' Classifier and Also IP addresses of the sender are often detected .

Introduction

In recent years, internet has become an integral part of life. With increased use of internet, numbers of email users are increasing day by day. This increasing use of email has created problems caused by unsolicited bulk email messages commonly referred to as Spam. Email has now become one of the best ways for advertisements due to which spam emails are generated. Spam emails are the emails that the receiver does not wish to receive. a large number of identical messages are sent to several recipients of email. Spam usually arises as a result of giving out our email address on an unauthorized or unscrupulous website .There

are many of the effects of Spam .Fills our Inbox with number of ridiculous emails .Degrades our Internet speed to a great extent .Steals useful information like our details on you Contact list .Alters your search results on any computer program .Spam is a huge waste of everybody's time and can quickly become very frustrating if you receive large amounts of it .Identifying these spammers and the spam content is a laborious task . even though extensive number of studies have been done, yet so far the methods set forth still scarcely distinguish spam surveys, and none of them demonstrate the benefits of each removed element compose . In spite of increasing network communication and wasting lot of memory space ,spam messages are also used for some attack . Spam emails, also known as non-self, are unsolicited commercial or malicious emails, sent to affect either a single individual or a corporation or a bunch of people. Besides advertising, these may contain links to phishing or malware hosting websites found out to steal confidential information. to solve this problem the different spam filtering techniques are used. spam filtering techniques are accustomed protect our mailbox for spam mails.

<u>Against this backdrop, this research paper sets out to achieve several objectives:</u>

I. Data Collection: Gather a large and diverse dataset of emails, including both spam and legitimate (ham) emails. This dataset should cover various types of spam emails, including phishing, scam, promotional, and malware-containing emails.

II. Data Preprocessing: Clean and preprocess the email dataset to remove noise, such as HTML tags, special characters, and irrelevant metadata. Convert the emails into a format suitable for analysis, such as tokenizing the text and removing stop word

III. Feature Engineering: Extract relevant features from the preprocessed email data. This might include features such as word frequencies, n-grams, presence of certain keywords, email header information (e.g., sender, subject), and structural characteristics of the email (e.g., number of attachments, URL links).
IV. Labeling: Annotate the dataset with labels indicating whether each email is spam or ham. This labeling process is crucial for supervised learning algorithms to train accurate classification models.
V. Feedback Mechanism: Implement a feedback mechanism where users can report misclassified emails to continuously improve the spam detection system. Use this feedback to retrain the model periodically and adapt to evolving spam patterns.

Method

Dataset Description

Dataset Title**:** Email Spam Classification Dataset

Dataset Source:

https://www.kaggle.com/datasets/balaka18/email-spam-classification-dataset-csv

Description:

A spam email classification dataset is a collection of labeled emails used for training, validating, and testing machine learning models to distinguish between spam (unsolicited or unwanted emails) and legitimate (desired or solicited emails). The dataset consists of a large number of individual email samples, each representing a single email message. These emails can vary in length, content, and format, reflecting the diversity of real-world email communication. Each email in the dataset is labeled as either spam or legitimate (ham). This labeling is crucial for supervised learning algorithms to learn the

patterns and characteristics that distinguish spam from legitimate emails.

Exploratory Data Analysis And Data Visualization

The database consist of 4373 rows and 2 columns.

The key attributes included in dataset are:

- Text Content
- Label
- Spam
- Ham

Spam Detection Using Machine Learning:

1. For training the algorithm dataset from Kaggle is used which is shown below

```
import pandas as pd
import numpy as np
```

```
data=pd.read_csv("spam.csv", encoding="latin-1")
```

```
data.head()
```

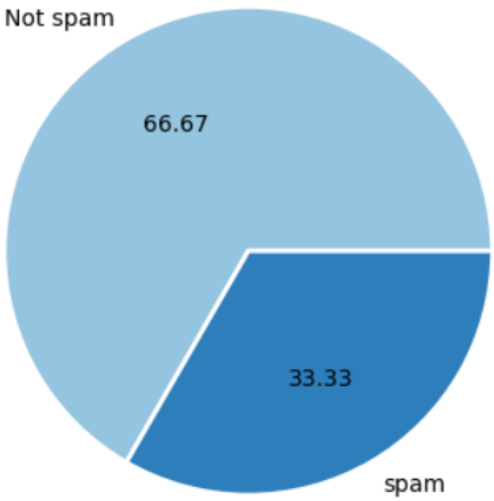
	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN

2. A pie chart to visualize the distribution of values in that column:

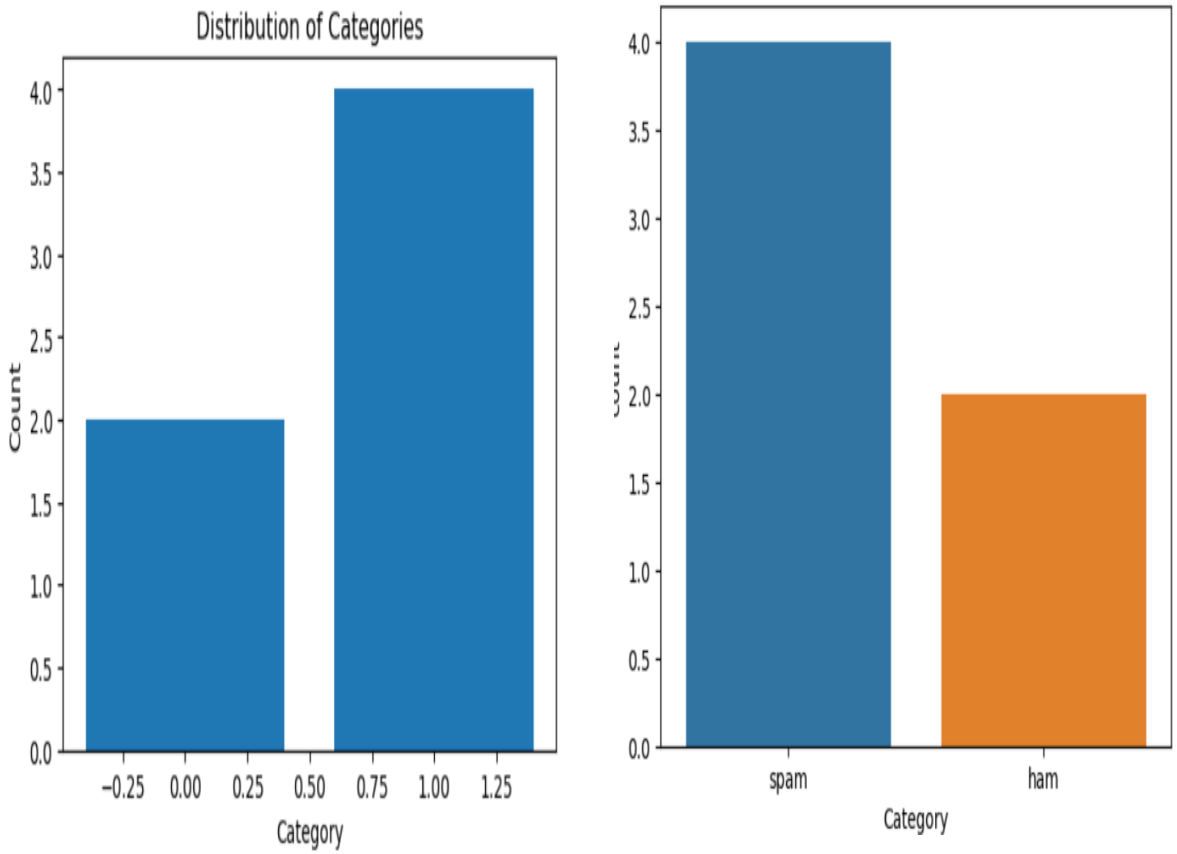
[59]:
```
# Now you can access the 'Category' column
plot_spam_ratio = df['Category'].value_counts().rename_axis('values').to_frame('counts').reset_index()

# Now you can plot the pie chart
import plotly.express as px
px.pie(plot_spam_ratio, values='counts', names='values')
```


```
from sklearn.preprocessing import LabelEncoder
encoder = LabelEncoder()
```

```
import matplotlib.pyplot as plt
colors = plt.get_cmap('Blues')(np.linspace(0.4, 0.7, 2))
plt.pie(data['Category'].value_counts(),labels=['Not spam','spam'],autopct='%0.2f', colors=colors,
        wedgeprops={"linewidth": 2, "edgecolor": "white"})
plt.show()
```

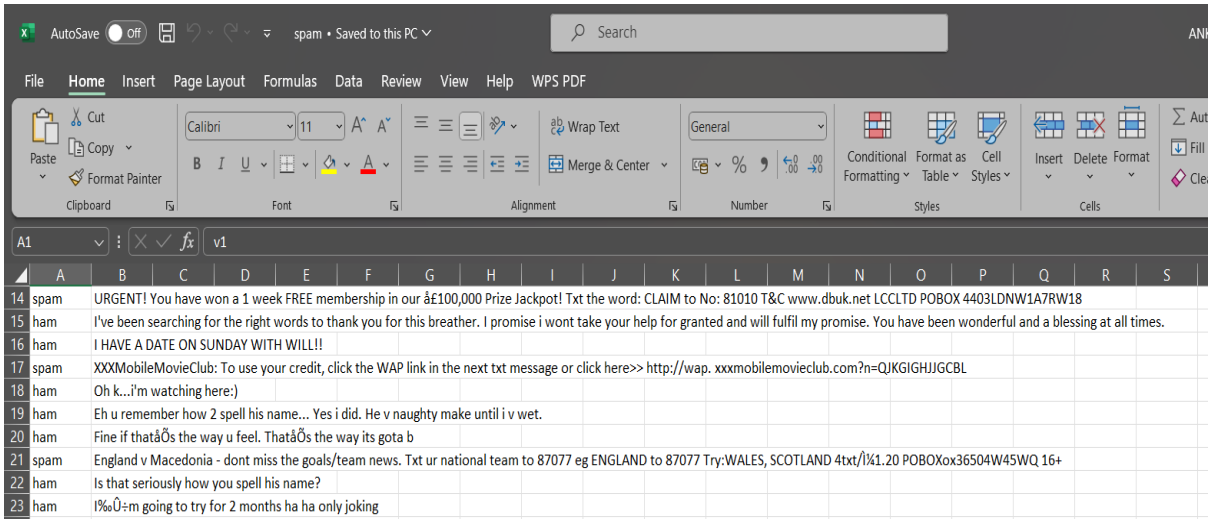


3.A bar plot to visualize the distribution of categories in the category:



4.Results and Discussions:

The exported message will be detected as spam or not using Bayes' theorem and Naive Bayes' Classifier.The exported dataset is:



	A	B
14	spam	URGENT! You have won a 1 week FREE membership in our å£100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C www.dbuk.net LCCLTD POBOX 4403LDNW1A7RW18
15	ham	I've been searching for the right words to thank you for this breather. I promise i wont take your help for granted and will fulfil my promise. You have been wonderful and a blessing at all times.
16	ham	I HAVE A DATE ON SUNDAY WITH WILL!!
17	spam	XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> http://wap. xxxmobilemovieclub.com?n=QJKGIGHJJGCBL
18	ham	Oh k...i'm watching here:)
19	ham	Eh u remember how 2 spell his name... Yes i did. He v naughty make until i v wet.
20	ham	Fine if thatåÕs the way u feel. ThatåÕs the way its gota b
21	spam	England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WALES, SCOTLAND 4txt/Ì¼1.20 POBOXox36504W45WQ 16+
22	ham	Is that seriously how you spell his name?
23	ham	I‰Û÷m going to try for 2 months ha ha only joking

The message is detected as Spam as shown below:

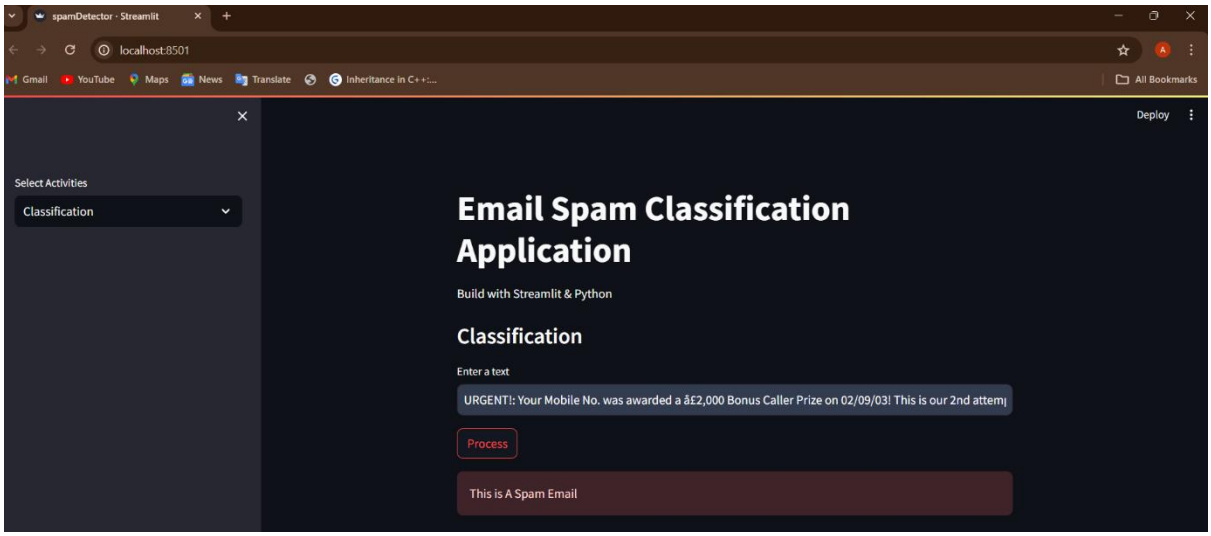
File Edit Selection View Go Run Terminal Help

spamDetector.py 3

C: > Users > wagav > Desktop > Mini project > spamDetector.py > ...

```
15
16    def main():
17        st.title("Email Spam Classification Application")
18        st.write("Build with Streamlit & Python")
19        activites=["Classification","About"]
20        choices=st.sidebar.selectbox("Select Activities",activites)
21        if choices=="Classification":
22            st.subheader("Classification")
23            msg=st.text_input("Enter a text")
24            if st.button("Process"):
25                print(msg)
26                print(type(msg))
27                data=[msg]
28                print(data)
29                vec=cv.transform(data).toarray()
30                result=model.predict(vec)
31                if result[0]==0:
32                    st.success("This is Not A Spam Email")
33                    speak("This is Not A Spam Email")
34                else:
35                    st.error("This is A Spam Email")
36                    speak("This is A Spam Email")
37    main()
```

Result:



Key Findings:

1.Spam Detection Accuracy:

The accuracy of the spam detection model is crucial. Key findings would include the overall accuracy of the model on the test dataset, indicating how well it distinguishes between spam and non-spam emails.

Identifying the most important features or words that contribute to identifying an email as spam. This could include specific keywords, phrases, sender information, email structure, or metadata.

2.Email Length Distribution:

Examining the distribution of email lengths (in terms of word count or character count) for spam and non-spam emails can highlight differences in writing style and content structure. For example, spam emails may tend to be shorter and more concise, while non-spam emails may contain longer, more elaborate messages.

3.Voice Message:

Input: The Email has been sent in the form of the text message by the sender Output: The email has been read through the use of voice note by the receiver.

4.Email Body Content:

Analyzing the presence of specific keywords or phrases in the email body that are indicative of spam (e.g., "discount," "free," "act now"). Certain words or patterns may be more prevalent in spam emails compared to legitimate emails. Examining common words or phrases in email subject lines for spam and non-spam emails. Spam emails often use attention-grabbing or misleading subject lines to entice recipients to open them.

Implications:

1.User Trust and Satisfaction:

By minimizing the presence of unwanted and malicious emails in users' inboxes, spam prediction projects enhance user trust and satisfaction. Users are more likely to engage with email platforms that provide a secure and pleasant user experience, fostering loyalty and retention.

2.Consumer Protection:

Spam prediction projects play a crucial role in protecting consumers from fraudulent schemes, identity theft, and online scams perpetrated through spam emails. By identifying and blocking malicious content, these projects contribute to consumer safety in the digital realm.

3.Cybersecurity Enhancement:

Effective spam prediction helps bolster cybersecurity defenses by identifying and filtering out potentially harmful emails containing phishing scams, malware, or fraudulent content. This reduces the risk of users falling victim to cyberattacks and protects sensitive information.

4.Algorithm Optimization and Operational Efficiency:

Spam prediction algorithms optimize resource allocation by reducing the burden on email servers, network bandwidth, and storage infrastructure. This improves the operational efficiency of email systems and lowers associated costs for organizations.

Future Research:

1.Adversarial Attacks:

Investigate techniques to make spam prediction models more resilient to adversarial attacks. Adversaries may attempt to evade detection by crafting spam emails specifically designed to fool machine learning algorithms. Research into adversarial training methods and robust feature representations could mitigate this risk.

2.Multi-modal Spam Detection:

Investigate the integration of multiple data modalities, such as text, images, and audio, for

more comprehensive spam detection. Incorporating information from email attachments, embedded images, and audio content could improve the accuracy of spam prediction models.

3.Explainable AI:

Develop techniques for making spam prediction models more interpretable and transparent. Explainable AI methods could help users understand why a particular email was classified as spam or ham, increasing trust in the prediction system and facilitating error analysis.

4.Online Learning and Adaptation:

Investigate online learning algorithms that can continuously adapt to new spamming techniques and evolving email content. Online learning approaches allow spam prediction models to learn from incoming email streams in real-time and update their predictions accordingly..

5.Privacy-preserving Techniques:

Develop privacy-preserving techniques for spam prediction that protect users' sensitive information while still maintaining high detection accuracy. Federated learning, differential privacy, and encrypted computation methods could be explored to address privacy concerns in spam detection.

Conclusion

Email has been the most important medium of communication nowadays, through internet connectivity any message can be delivered to all aver the world. More than 270 billion emails are exchanged daily, about 57% of these are just spam emails. Spam emails, also known as non-self, are undesired commercial or malicious emails, which affects or hacks personal information like bank ,related to money or anything that causes destruction to single individual or a corporation or a group of people. Besides advertising,

these may contain links to phishing or malware hosting websites set up to steal confidential information. Spam is a serious issue that is not just annoying to the end-users but also financially damaging and a security risk. Hence this system is designed in such a way that it detects unsolicited and unwanted emails and prevents them hence helping in reducing the spam message which would be of great benefit to individuals as well as to the company .In the future this system can be implemented by using different algorithms and also more features can be added to the existing system

References:

https://pandas.pydata.org/

https://www.kaggle.com/

https://matplotlib.org/

https://seaborn.pydata.org/