BUILDING A SMARTER AI-POWERED SPAM CLASSIFIER

The majority of people in today's society own a mobile phone, and they all frequently get communications (SMS/email) on their phones. But the key point is that some of the messages you get may be spam, with very few being genuine or important interactions. You may be tricked into providing your personal information, such as your password, account number, or Social Security number, by scammers that send out phony text messages. They may be able to access your bank, email, and other accounts if they obtain this information. To filter out these messages, a spam filtering system is used that marks a message spam on the basis of its contents or sender.

In this article, we will be seeing how to develop a spam classification system and also evaluate our model using various metrics. In this article, we will be majorly focusing on OpenAI API.

Abstract:- Spam Classification using Artificial Intelligence – For business purposes, email is the most widely utilized mode of official communication. Despite the availability of other forms of communication, email usage continues to rise. In today's world, automated email management is critical since the volume of emails grows by the day. More than 55

percent of all emails have been recognized as spam. This demonstrates that spammers waste email users' time and resources while producing no meaningful results. Spammers employ sophisticated and inventive strategies to carry out their criminal actions via spam emails.

As a result, it is critical to comprehend the many spam email classification tactics and mechanisms. The main focus of this paper is on spam classification using machine learning algorithms. Furthermore, this research includes a thorough examination and evaluation of research on several machine learning methodologies and email properties used in various Machine Learning approaches.

Future study goals and obstacles in the subject of spam classification are also discussed, which may be valuable to future researchers.

Objective: –

Machine learning algorithms use statistical models to classify data. In the case of spam detection, a trained machine learning model must be able to determine whether the sequence of words found in an email is closer to those found in spam emails or safe ones.

Introduction: –

For the majority of internet users, email has become the most often utilized formal communication channel. In recent years, there has been a surge in email usage, which has exacerbated the problems presented by spam emails. Spam, often known as junk email, is the act of sending unsolicited mass messages to a large number of people. ‘Ham’ refers to emails that are meaningful but of a different type. Every day, the average email user receives roughly 40-50 emails. Spammers earn roughly 3.5 million dollars per year from spam, resulting in financial damages on both a personal and institutional level. As a result, consumers devote a large amount of their working time to these emails. Spam is said to account for more than half of all email server traffic, sending out a vast volume of undesired and uninvited bulk emails.

They squander user resources on useless output, lowering productivity. Spammers use spam for marketing goals to spread malicious criminal acts such as identity theft, financial disruptions, stealing sensitive information, and reputational damage.

The existing model of the system: –

Spam refers to the term, which is related to undesired content with low-quality information, Spam referred to the major drawback of mobile business. When comes to spam detection in the campus network they did the analysis using Incremental Learning. For Collecting Spam detection on web pages. Moreover Sending out a Spam message was also analyzed. Data Collection was done privately by a limited company. From the data Collection. There also anti-spam filter system was evolved. Many parallel and distributed computing system has also processed this spam system. Machine learning algorithm provides accurate result. Text Mining analysis done separates ham and spam separately.

Proposed model of the system: –

As we look at spam detection systems that use Machine Learning (ML) techniques, it’s vital to take a look at the history of ML in the field as well as the many methods that are now used to identify spam. Researchers have discovered that the content of spam emails, as well as their operational procedures, evolve with time.

As a result, the tactics that are currently effective may become obsolete in the near future. The conceptual drift [8] is a term used to describe this occurrence.

Machine Learning is an engineering approach that allows computational instruments to behave without being explicitly programmed. Because of the ML system’s ability to evolve, limiting concept drift, this strategy is a significant help in detecting and combating spam.

In the next section, we'll go through a variety of machine learning techniques, approaches, and algorithms, as well as the benefits of each, using Supervised, Unsupervised, and Semi-Supervised Machine Learning algorithms Approaches.

System Requirements: –

Hardware

1.OS – Windows 7, 8, and 10 (32 and 64 bit)

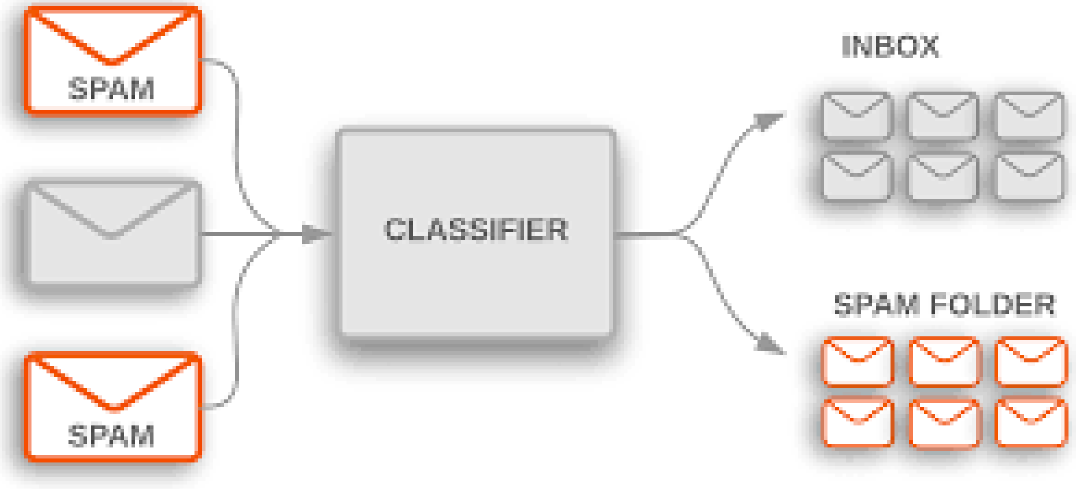
2.RAM – 4GB

Software:

1.Python

2.Anaconda navigator

3.Python built-in module

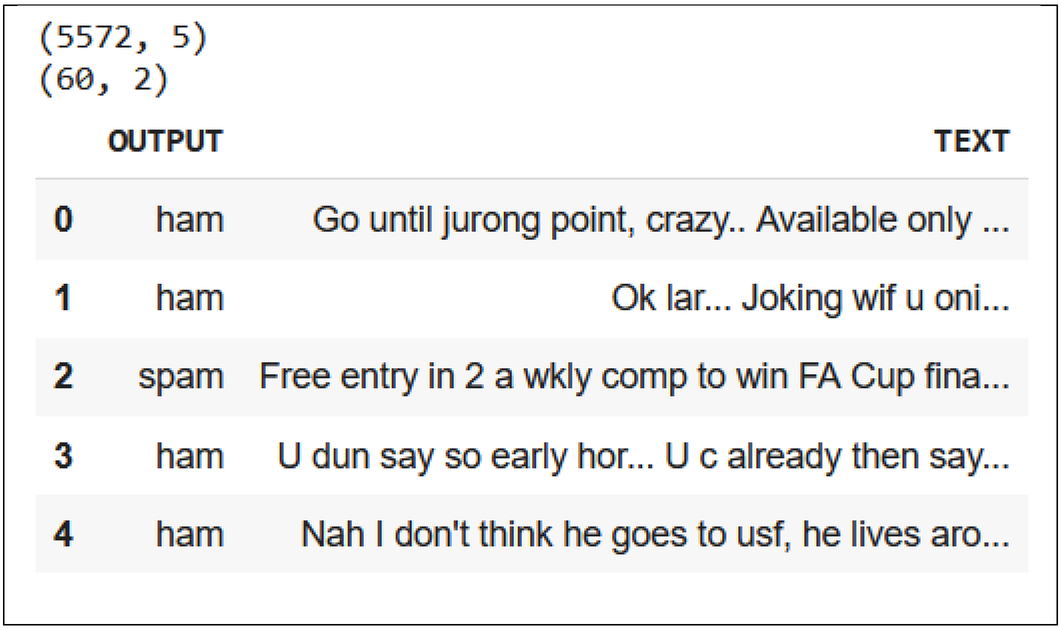


PROGRAM:-

```
# necessary libraries
import openai
import pandas as pd
import numpy as np
# libraries to develop and evaluate a machine learning model
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
accuracy_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
accuracy_score
from sklearn.metrics import confusion_matrix
# replace "YOUR API KEY" with your generated API key
openai.api_key = "YOUR API KEY"
# while loading the csv, we ignore any encoding errors and
skip any bad line
df = pd.read_csv('spam.csv', encoding_errors='ignore',
on_bad_lines='skip')
print(df.shape)
```

```
# we have 3 columns with NULL values, to remove that we use the below line

df = df.dropna(axis=1)

# we are taking only the first 60 rows for developing the model

df = df.iloc[:60]

# rename the columns v1 and v2 to Output and Text respectively

df.rename(columns = {'v1':'OUTPUT', 'v2': 'TEXT'}, inplace = True)

print(df.shape)

df.head()
```

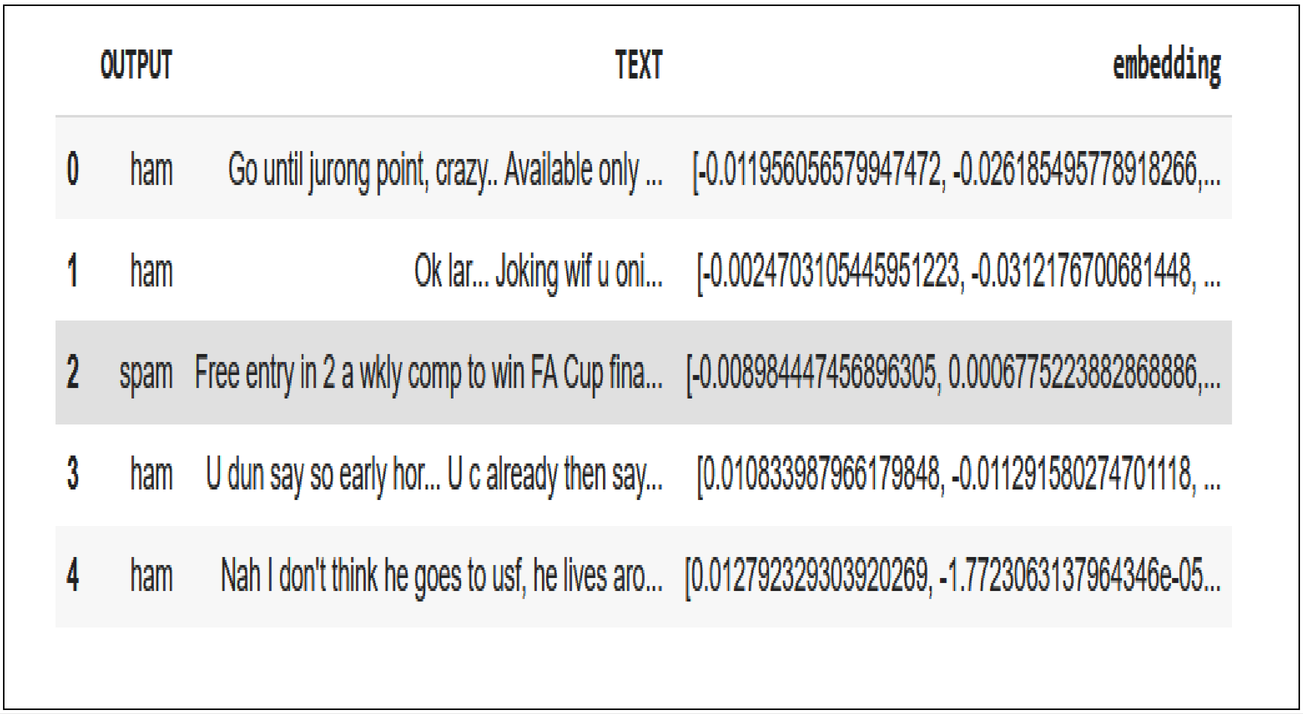
OUTPUT:



```
(5572, 5)
(60, 2)
     OUTPUT                                             TEXT
0       ham    Go until jurong point, crazy.. Available only ...
1       ham                        Ok lar... Joking wif u oni...
2      spam   Free entry in 2 a wkly comp to win FA Cup fina...
3       ham     U dun say so early hor... U c already then say...
4       ham     Nah I don't think he goes to usf, he lives aro...
```

```
# function to generate vector for a string
def get_embedding(text, model="text-embedding-ada-002"):
   return openai.Embedding.create(input = ,
model=model)['data'][0]['embedding']


# applying the above funtion to generate vectors for all 60
text pieces
df["embedding"] =
df.TEXT.apply(get_embedding).apply(np.array)  # convert
string to array
df.head()
```

OUTPUT:



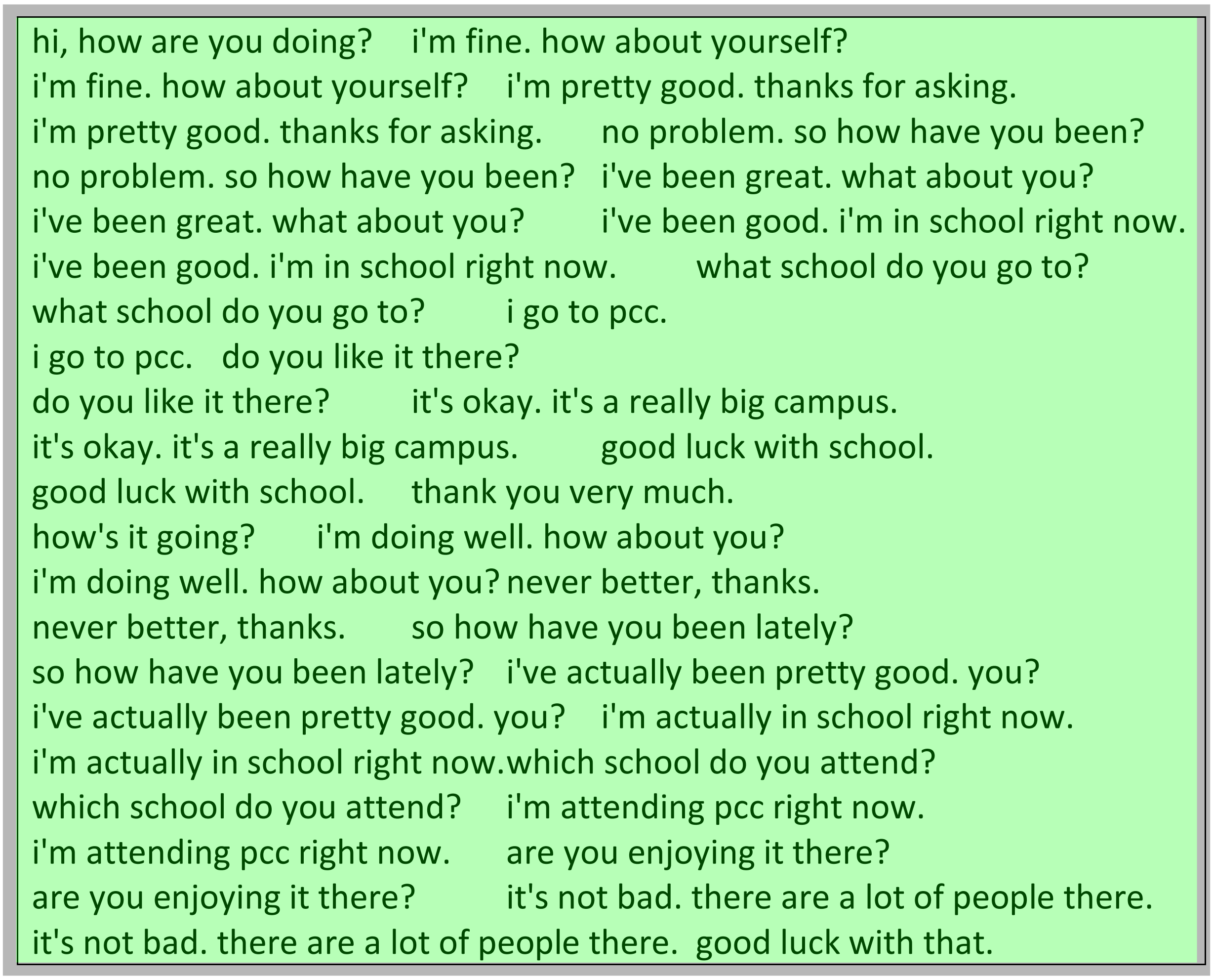
	OUTPUT	TEXT	embedding
0	ham	Go until jurong point, crazy.. Available only ...	[-0.011956056579947472, -0.026185495778918266,...
1	ham	Ok lar... Joking wif u oni...	[-0.0024703105445951223, -0.0312176700681448, ...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	[-0.008984447456896305, 0.000677522388286886,...
3	ham	U dun say so early hor... U c already then say...	[0.010833987966179848, -0.011291580274701118, ...
4	ham	Nah I don't think he goes to usf, he lives aro...	[0.012792329303920269, -1.7723063137964346e-05...

```
'# split data into train and test
X = np.array(df.embedding)
y = np.array(df.class_embeddings)
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)


# train random forest classifier
clf = RandomForestClassifier(n_estimators=100)
clf.fit(X_train.tolist(), y_train)
preds = clf.predict(X_test.tolist())
   # generate a classification report involving f1-score, recall,
precision and accuracy
report = classification_report(y_test, preds)
print(report)
```

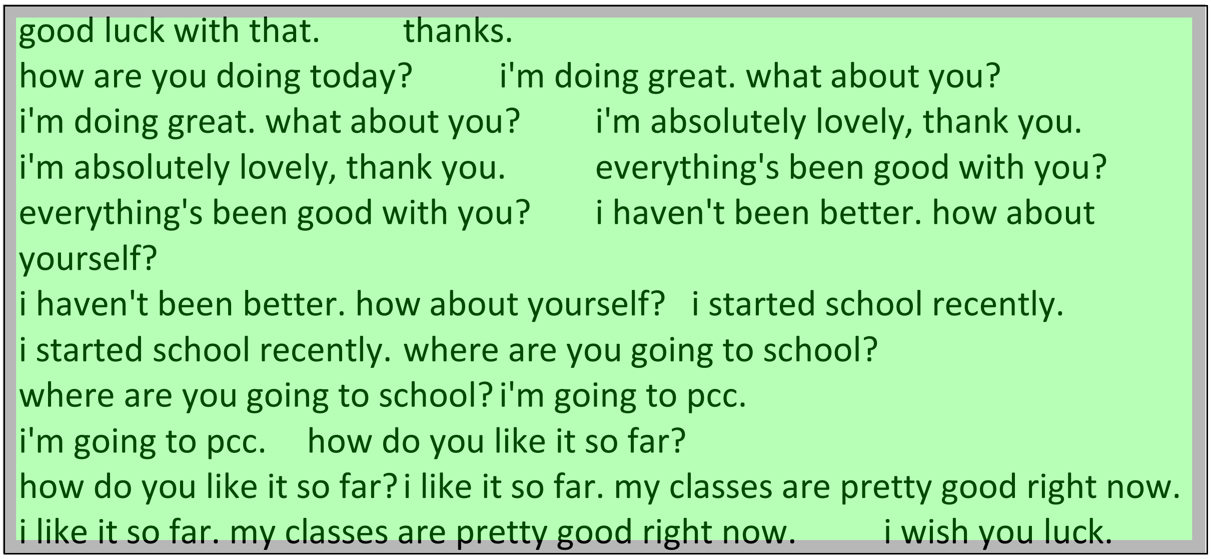
OUTPUT:

```
                      precision    recall      f1-score    support

       0                 0.82       1.00         0.90          9

       1                1.00        0.33         0.50         3

   accuracy                                      0.83         12

  macro avg            0.91       0.67         0.70          12

weighted avg            0.86      0.83         0.80          12
```

DATA SET FOR AI-POWERED SPAM CLASSIFIER



```
hi, how are you doing?    i'm fine. how about yourself?
i'm fine. how about yourself?    i'm pretty good. thanks for asking.
i'm pretty good. thanks for asking.       no problem. so how have you been?
no problem. so how have you been?   i've been great. what about you?
i've been great. what about you?        i've been good. i'm in school right now.
i've been good. i'm in school right now.         what school do you go to?
what school do you go to?         i go to pcc.
i go to pcc.    do you like it there?
do you like it there?         it's okay. it's a really big campus.
it's okay. it's a really big campus.         good luck with school.
good luck with school.     thank you very much.
how's it going?      i'm doing well. how about you?
i'm doing well. how about you? never better, thanks.
never better, thanks.       so how have you been lately?
so how have you been lately?    i've actually been pretty good. you?
i've actually been pretty good. you?    i'm actually in school right now.
i'm actually in school right now.which school do you attend?
which school do you attend?     i'm attending pcc right now.
i'm attending pcc right now.      are you enjoying it there?
are you enjoying it there?          it's not bad. there are a lot of people there.
it's not bad. there are a lot of people there.  good luck with that.
```



```
good luck with that.	thanks.
how are you doing today?	i'm doing great. what about you?
i'm doing great. what about you?	i'm absolutely lovely, thank you.
i'm absolutely lovely, thank you.	everything's been good with you?
everything's been good with you?	i haven't been better. how about yourself?
i haven't been better. how about yourself?	i started school recently.
i started school recently.	where are you going to school?
where are you going to school?	i'm going to pcc.
i'm going to pcc.	how do you like it so far?
how do you like it so far?	i like it so far. my classes are pretty good right now.
i like it so far. my classes are pretty good right now.	i wish you luck.
```

Conclusion: –

Following a thorough examination of the chosen study, Several study findings and observations have been identified as a result of our studies. These were previously discussed in detail.

portions that are well-explained In this section, we'll talk about concentrating more on the major findings and conclusions of the research Supervised machine learning has a high acceptance rate. Throughout the review, the approach can be noticed. This strategy is effective. is employed primarily because it produces more accurate findings. With less fluctuation, this strategy has a high level of consistency. Aside from that, we've discovered that certain algorithms work better than others. When compared to other techniques, such as Nave Based and SVM, there is a strong demand for them. Machine Learning Algorithms that aren't as well-known. The employed multi-algorithm. n order to achieve a better result, systems are increasingly commonly used. rather than a single algorithm

References:-

[1] “Global spam volume as a percentage of total e-mail traffic from January 2014 to September 2019, by month.”https://www.statista.com/statistics/420391/spam-email-traffific-share/.

[2] T. Ouyang, S. Ray, M. Allman, and M. Rabinovich, “A large-scale empirical analysis of email spam detection through network characteristics in a stand-alone enterprise,” Elsevier, vol. 2015, pp. 101–102.

THANK YOU!!!