

BUILDING A SMARTER AI-POWERED SPAM CLASSIFIER

REQUIREMENTS:

Hardware: PC with Core-I7.

Software : Windows 11, SPYDER.

ALGORITHM:

There are many types of algorithms used in spam classification like., Naïve bayes, Multilayer Perceptron, Decision Tree Etc.,

In this project, We use Naive Bayes algorithm.

Naive:

It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.

Bayes Algorithm:

It is called Bayes because it depends on the principle of [Bayes' Theorem](#).

- Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

- **P(A|B) is Posterior probability:** Probability of hypothesis A on the observed event B.
- **P(B|A) is Likelihood probability:** Probability of the evidence given that the probability of a hypothesis is true.
- **P(A) is Prior Probability:** Probability of hypothesis before observing the evidence.
- **P(B) is Marginal Probability:** Probability of Evidence

Dataset Link: <https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset>

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, 0.0006775223882868886, ...
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) # 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			

```
print("accuracy: ", np.round(accuracy_score(y_test, preds)*100,2), "%")
```

OUTPUT:

Accuracy: 83.33%

```
confusion_matrix(y_test, preds)
```

OUTPUT:

```
Array([[9,0],[2,1]])
```

```
import openai

# replace "YOUR API KEY" with your generated API key
openai.api_key = "YOUR API KEY"

def spam_classification(message):
    response = openai.Completion.create(
        model="text-davinci-003",
        prompt=f"Classify the following message as spam or not\nspam:\n\n{message}\n\nAnswer:",
        temperature=0,
        max_tokens=64,
        top_p=1.0,
        frequency_penalty=0.0,
        presence_penalty=0.0
    )
    return response['choices'][0]['text'].strip()

out = spam_classification("""Congratulations! You've Won a $1000 gift card from walmart.

Go to https://bit.ly to claim your reward.""")

print(out)
```

OUTPUT:

Spam

```
out = spam_classification("Hey Alex, just wanted to let you know  
tomorrow is an off. Thank you")
```

```
print(out)
```

OUTPUT:

Not a spam

Data Set:

Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore w
Ok lar... Joking wif u oni...

Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry c
U dun say so early hor... U c already then say...

Nah I don't think he goes to usf, he lives around here though

FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it stil
Even my brother is not like to speak with me. They treat me like aids patent.

As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your caller
WINNER!! As a valued network customer you have been selected to receive a £900 prize reward! To c
Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera
I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough
SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6
URGENT! You have won a 1 week FREE membership in our £100,000 Prize Jackpot! Txt the word: CLA
I've been searching for the right words to thank you for this breather. I promise i wont take your help f
I HAVE A DATE ON SUNDAY WITH WILL!!

XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> htt

Oh k...i'm watching here:)

EH u remember how 2 spell his name... Yes i did. He v naughty make until i v wet.

Fine if that's the way u feel. That's the way its gota b

England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to
Is that seriously how you spell his name?

I'm going to try for 2 months ha ha only joking

So ò_ pay first lar... Then when is da stock comin...

Aft i finish my lunch then i go str down lor. Ard 3 smth lor. U finish ur lunch already?

Ffffffffff. Alright no way I can meet up with you sooner?

Just forced myself to eat a slice. I'm really not hungry tho. This sucks. Mark is getting worried. He know
Lol your always so convincing.

Did you catch the bus ? Are you frying an egg ? Did you make a tea? Are you eating your mom's left ov
I'm back & we're packing the car now, I'll let you know if there's room

Ahhh. Work. I vaguely remember that! What does it feel like? Lol

Wait that's still not all that clear, were you not sure about me being sarcastic or that that's why x does

OUTPUT:

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