

## Experiment No : 2

**Title:-**Write a program to detecting YouTube comment spam.

```
In [16]: import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import BernoulliNB
```

```
In [17]: data = pd.read_csv("Youtube01.csv")
```

```
In [18]: data.head()
```

Out[18]:

	COMMENT_ID	AUTHOR	DATE	CONTENT	CLASS
0	LZQPQhLyRh80UYxNuaDWhIGQYNQ96luCg-AYWqNPjpU	Julius NM	2013-11-07T06:20:48	Huh, anyway check out this you[tube] channel: ...	1
1	LZQPQhLyRh_C2cTtd9MvFRJedxydaVW-2sNg5Diuo4A	adam riyati	2013-11-07T12:37:15	Hey guys check out my new channel and our firs...	1
2	LZQPQhLyRh9MSZYnf8djyK0gEF9BHDPYrrK-qCczIY8	Evgeny Murashkin	2013-11-08T17:34:21	just for test I have to say murdev.com	1
3	z13jhp0bxqncu512g22wvzkasxmvvzjaz04	ElNino Melendez	2013-11-09T08:28:43	me shaking my sexy ass on my channel enjoy ^ ^ _	1
4	z13fwbwp1oujthgqj04chlngpvzmtt3r3dw	GsMega	2013-11-10T16:05:38	watch?v=vtaRGgvGtWQ Check this out .	1

```
In [19]: # Since we only require content and class columns, we will update our exist
ing dataframe to the following
data = data[['CONTENT', 'CLASS']]

# We will map 0 to not spam and 1 to spam in class column
data["CLASS"] = data['CLASS'].map({0: 'NOT A SPAM COMMENT', 1: 'SPAM COMMENT'})
print(data.sample(5))

x = np.array(data['CONTENT'])
y = np.array(data['CLASS'])

"""
As the output of this problem will either be 0 or 1, i.e., the problem of binary classification,
we can use the Bernoulli Naive Bayes algorithm to train the model:
"""

cv = CountVectorizer()
x = cv.fit_transform(x)
xtrain, xtest, ytrain, ytest = train_test_split(x, y, train_size=0.2, random_state=42)

model = BernoulliNB()
model.fit(xtrain, ytrain)

print(model.score(xtest, ytest))

# Its time to check the model by giving spam and non-spam comments

S = input("Enter a comment : ")
d = cv.transform([S]).toarray()
print(model.predict(d))
```

	CONTENT	CLASS
637	Oh my god ... Roar is the most liked video at ...	NOT A SPAM COMMENT
438	The great mother of the jungle. Sweet and natu...	NOT A SPAM COMMENT
1955	Shakira is the best dancer	NOT A SPAM COMMENT
1241	MEGAN FOX AND EMINEM TOGETHER IN A VIDEO DOES...	NOT A SPAM COMMENT
1776	visit &quot; ww estiloproduction com &quot; be...	SPAM COMMENT

0.8223642172523962

Enter a comment : I think this is now a place to promote channels in the comment section lol.

['NOT A SPAM COMMENT']

In [ ]: