

```
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
data=pd.read_csv("Youtube01-Psy.csv")
print(data.head())
```

	COMMENT_ID	AUTHOR
0	LZQPQhLyRh80UYxNuaDWhIGQYNQ96IuCg-AYWqNPjpU	Julius NM
1	LZQPQhLyRh_C2cTtd9MvFRJedxydaVW-2sNg5Divo4A	adam riyati
2	LZQPQhLyRh9MSZYnf8djyK0gEF9BHDPYrrK-qCczIY8	Evgeny Murashkin
3	z13jhp0bxqncu512g22wvzkasxmvvzjaz04	ElNino Melendez
4	z13fwbwp1oujthgqj04chlngpvzmtt3r3dw	GsMega

	DATE	
0	2013-11-07T06:20:48	Huh, anyway check out this you[tube] chann
1	2013-11-07T12:37:15	Hey guys check out my new channel and our
2	2013-11-08T17:34:21	just for test I have to say mur
3	2013-11-09T08:28:43	me shaking my sexy ass on my channel enjo
4	2013-11-10T16:05:38	watch?v=vtaRGgvGtWQ Check this

	CLASS
0	1
1	1
2	1
3	1

```
data=data[["CONTENT", "CLASS"]]
print(data.sample(5))
```

	CONTENT	CLASS
135	Ahhh, 2 years ago....	0
105	Sub to my channel visuelgamingzNL I sub back	1
98	Anybody who subscribes to me will get 10 subsc...	1
97	Please help me go here http://www.gofundme.com...	1
201	i am 2,126,492,636 viewer :D	0

```
data["CLASS"] = data["CLASS"].map({0:"not spam",
                                     1:"spam"})
print(data.sample(5))
```

	CONTENT	CLASS
161	Incmedia.org where the truth meets you.	spam
143	Add me here...https://www.facebook.com/TLouXmu...	spam
253	I am now going to voyage to the first comment....	not spam
75	if your like drones, plz subscribe to Kamal Ta...	spam
94	Why does this video have so many views? Becaus...	not spam

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

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

model = BernoulliNB()
model.fit(xtrain, ytrain)
print(model.score(xtest, ytest))
```

0.9857142857142858

```
sample = "Check this out: https://thecleverprogrammer.com/"
data = cv.transform([sample]).toarray()
print(model.predict(data))
```

['not spam']

```
sample = "Lack of information!"
data = cv.transform([sample]).toarray()
print(model.predict(data))
```

['not spam']