

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import CountVectorizer
```

```
dataframe=pd.read_csv("spam.csv",encoding='latin-1')
dataframe=dataframe[["v1","v2"]]
dataframe
```

```

      v1                                                    v2
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...
...     ...
5567  spam  This is the 2nd time we have tried 2 contact u...
5568  ham                                Will I_b going to esplanade fr home?
5569  ham  Pity, * was in mood for that. So...any other s...
5570  ham  The guy did some bitching but I acted like i'd...
5571  ham                                Rofl. Its true to its name
```

```
[5572 rows x 2 columns]
```

```
dataframe.rename(columns={"v1":"Category","v2":"Message"},inplace=True)
dataframe.head()
```

```

      Category                                                    Message
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...
```

```
dataframe.groupby("Category").describe()
```

```

      Message
\
      count unique
top
Category
ham      4824    4510      Sorry, I'll call
later
spam      747     654  Please call our customer service
representativ...
```

```
freq
```

```
Category
ham      30
spam      4
```

```
dataframe["spam"]=dataframe["Category"].apply(lambda x:1 if x=="spam"
else 0)
```

```
dataframe
```

	Category	Message	spam
0	ham	Go until jurong point, crazy.. Available only ...	0
1	ham	Ok lar... Joking wif u oni...	0
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	1
3	ham	U dun say so early hor... U c already then say...	0
4	ham	Nah I don't think he goes to usf, he lives aro...	0
...
5567	spam	This is the 2nd time we have tried 2 contact u...	1
5568	ham	Will I_b going to esplanade fr home?	0
5569	ham	Pity, * was in mood for that. So...any other s...	0
5570	ham	The guy did some bitching but I acted like i'd...	0
5571	ham	Rofl. Its true to its name	0

```
[5572 rows x 3 columns]
```

```
independent_var=dataframe["Message"]
```

```
dependent_var=dataframe["spam"]
```

```
x_train,x_test,y_train,y_test=train_test_split(independent_var,depende
nt_var,train_size=0.75,random_state=43)
```

```
print("shape of x train data is",x_train.shape)
```

```
print("shape of y train data is",y_train.shape)
```

```
print("shape of x test data is",x_test.shape)
```

```
print("shape of y test data is",y_test.shape)
```

```
shape of x train data is (4179,)
```

```
shape of y train data is (4179,)
```

```
shape of x test data is (1393,)
```

```
shape of y test data is (1393,)
```

```
countvectobj=CountVectorizer()
```

```
x_train_count=countvectobj.fit_transform(x_train.values.astype('U'))
```

```
print(x_train_count.toarray())
```

```
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
```

```
modelobj=MultinomialNB()
modelobj.fit(x_train_count,y_train)

MultinomialNB()

email=[
    'Can you send the documents',
    'Sale!! get 50% off'
]
email_count=countvectobj.transform(email)
modelobj.predict(email_count)

array([0, 1])

x_test_count=countvectobj.transform(x_test)
modelobj.score(x_test_count,y_test)

0.9827709978463748
```