emailspamprediction

October 16, 2023

[50]:
```
import pandas as pd
import numpy as np
import seaborn as sns
```

[90]:
```
df = pd.read_csv('emails.csv')
df
```

[90]:

	text	spam
0	Subject: naturally irresistible your corporate…	1
1	Subject: the stock trading gunslinger fanny i…	1
2	Subject: unbelievable new homes made easy im …	1
3	Subject: 4 color printing special request add…	1
4	Subject: do not have money , get software cds …	1
…	…	…
5723	Subject: re : research and development charges…	0
5724	Subject: re : receipts from visit jim , than…	0
5725	Subject: re : enron case study update wow ! a…	0
5726	Subject: re : interest david , please , call…	0
5727	Subject: news : aurora 5 . 2 update aurora ve…	0

[5728 rows x 2 columns]

[70]:
```
df['spam'].value_counts()
```

[70]:
```
spam
0    4360
1    1368
Name: count, dtype: int64
```

[71]:
```
df.drop_duplicates(inplace = True)
```

[72]:
```
df['spam'].value_counts()
```

[72]:
```
spam
0    4327
1    1368
Name: count, dtype: int64
```

[73]:
```
df
```

[73]:
```
                                                   text  spam
0     Subject: naturally irresistible your corporate…     1
1     Subject: the stock trading gunslinger  fanny i…     1
2     Subject: unbelievable new homes made easy  im …     1
3     Subject: 4 color printing special  request add…     1
4     Subject: do not have money , get software cds …     1
…                                                   …   …
5723  Subject: re : research and development charges…     0
5724  Subject: re : receipts from visit  jim ,  than…     0
5725  Subject: re : enron case study update  wow ! a…     0
5726  Subject: re : interest  david ,  please , call…     0
5727  Subject: news : aurora 5 . 2 update  aurora ve…     0

[5695 rows x 2 columns]
```

[74]:
```
df.isnull().sum()
```

[74]:
```
text    0
spam    0
dtype: int64
```

[75]:
```
x = df.text.values
```

[76]:
```
y = df.spam.values
```

[77]:
```
from sklearn.model_selection import train_test_split
```

[78]:
```
xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size=0.2)
```

[79]:
```
from sklearn.feature_extraction.text import CountVectorizer
```

[80]:
```
cv = CountVectorizer()
x_train = cv.fit_transform(xtrain)
```

[81]:
```
x_train.toarray()
```

[81]:
```
array([[4, 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]], dtype=int64)
```

[82]:
```
from sklearn.naive_bayes import MultinomialNB
```

[83]:
```
model = MultinomialNB()
model.fit(x_train,ytrain)
```

[83]:
```
MultinomialNB()
```

[84]:
```
x_test = cv.transform(xtest)
```

[85]:
```
x_test.toarray()
```

[85]:
```
array([[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]], dtype=int64)
```

[86]:
```
model.score(x_test,ytest)
```

[86]:
```
0.9912203687445127
```

[89]:
```
emails = input()
emails = [emails]
cv_emails = cv.transform(emails)

if model.predict(cv_emails) == 1:
    print("This is Spam Email")
else:
    print("This is not a spam email")
```

```
Allagh is Almighty
This is not a spam email
```

[ ]:
```

```