ass6_mnish_spam

February 10, 2026

```python
[33]: import numpy as np
      import pandas as pd
```

```python
[35]: data=pd.read_csv('/home/csl-4/Downloads/spam.csv')
      data
```

```
[35]:      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…
      …         …                                                 …
      5567     spam  This is the 2nd time we have tried 2 contact u…
      5568      ham              Will ü 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]
```

```python
[37]: data.columns
```

```
[37]: Index(['Category', 'Message'], dtype='object')
```

```python
[39]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 2 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Category  5572 non-null   object
 1   Message   5572 non-null   object
dtypes: object(2)
memory usage: 87.2+ KB
```

```python
[41]: data['Spam']=data['Category'].apply(lambda x:1 if x=='spam' else 0)
      data.head(5)
```

```
[41]:    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
```

```python
[43]: from sklearn.model_selection import train_test_split
      X_train,X_test,y_train,y_test=train_test_split(data.Message,data.
        ↪Spam,test_size=0.75)
```

```python
[45]: from sklearn.feature_extraction.text import CountVectorizer
```

```python
[47]: from sklearn.naive_bayes import MultinomialNB
```

```python
[49]: from sklearn.pipeline import Pipeline
      clf=Pipeline([
          ('vectorizer',CountVectorizer()),
          ('nb',MultinomialNB())
      ])
```

```python
[51]: clf.fit(X_train,y_train)
```

```
[51]: Pipeline(steps=[('vectorizer', CountVectorizer()), ('nb', MultinomialNB())])
```

```python
[53]: emails=[
          'Sounds great! Are you home now?',
          'Will u meet ur dream partner soon? Is ur career off 2 a flyng start? 2␣
        ↪find out free, txt HORO followed by ur star sign, e. g. HORO ARIES'
      ]
```

```python
[55]: clf.predict(emails)
```

```
[55]: array([0, 1])
```

```python
[57]: clf.score(X_test,y_test)
```

```
[57]: 0.9775065805216558
```

```python
[ ]:
```

```python
[ ]:
```