



5572 rows x 5 columns

In [10]: `x = df.iloc[:, :-3]`

In [11]: `df`

Out[11]:

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN
...	...	...	...	...	...
5567	spam	This is the 2nd time we have tried 2 contact u...	NaN	NaN	NaN
5568	ham	Will i_b going to esplanade fr home?	NaN	NaN	NaN
5569	ham	Pity, * was in mood for that. So...any other s...	NaN	NaN	NaN
5570	ham	The guy did some bitching but I acted like i'd...	NaN	NaN	NaN
5571	ham	Rofl. Its true to its name	NaN	NaN	NaN

5572 rows x 5 columns

In [12]: `x`

Out[12]:

	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...

LP1\_NBA Lab Manual\_2022-23.p...Desktop/ML/ML\_LAB/Assign\_4 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML\_LAB/Assign\_4.ipynb

Logout

Jupyter Assign\_4 (unsaved changes)

FileEditViewInsertCellKernelHelpNot TrustedPython 3 (ipykernel)

In [12]: x

Out[12]:

	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

In [14]: df.groupby('v1').describe()

Out[14]:

	count		unique		v2		Unnamed: 2				Unnamed: 3				Unnamed: 4	
	count		unique		top	freq	count	unique	top	freq	count	unique	top	freq	count	unique
v1																
ham	4825	4516			Sorry, I'll call later	30	45	39	bt not his girfrnd... Goodnight... @"	3	10	9	GE	2	6	5 GNT:-)" 2
spam	747	653			Please call our customer service representativ...	4	5	4	PO Box 5249	2	2	1	MK17 92H. 450Ppw 16"	2	0	0 NaN NaN

In [16]: df['spam'] = df['v1'].apply(lambda x: 1 if x == 'spam' else 0)

df.head()

Out[16]:

LP1\_NBA Lab Manual\_2022-23.pptx

Desktop/ML/ML\_LAB/

Assign\_4 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML\_LAB/Assign\_4.ipynb

jupyter Assign\_4 (unsaved changes)

Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

In [16]:

df['spam'] = df['v1'].apply(lambda x: 1 if x == 'spam' else 0)  
df.head()

Out[16]:

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4	spam
0	ham Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN	NaN	0
1	ham Ok lar... Joking wif u oni...	NaN	NaN	NaN	NaN	0
2	spam Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN	NaN	1
3	ham U dun say so early hor... U c already then say...	NaN	NaN	NaN	NaN	0
4	ham Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN	NaN	0

In [18]:

from sklearn.model\_selection import train\_test\_split  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(df.v2, df.spam)

In [19]:

from sklearn.feature\_extraction.text import CountVectorizer  
v = CountVectorizer()  
X\_train\_count = v.fit\_transform(X\_train.values)  
X\_train\_count.toarray()[ :2]

Out[19]:

array([[0, 0, 0, ..., 0, 0, 0],  
 [0, 0, 0, ..., 0, 0, 0]])

In [20]:

from sklearn.naive\_bayes import MultinomialNB  
model = MultinomialNB()  
model.fit(X\_train\_count, y\_train)

Out[20]:

MultinomialNB()

In [21]:

emails = ['How are you brother?', 'Get Flat 50% on Your body outfits']  
emails\_count = v.transform(emails)  
model.predict(emails\_count)

Out[21]:

array([0, 1])

In [22]:

X\_test\_count = v.transform(X\_test)



File Edit View Insert Cell Kernel Help

Not Trusted

Python 3 (ipykernel) ○

```
In [18]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(df.v2, df.spam)
```

```
In [19]: from sklearn.feature_extraction.text import CountVectorizer
v = CountVectorizer()
X_train_count = v.fit_transform(X_train.values)
X_train_count.toarray()[:2]
```

```
Out[19]: array([[0, 0, 0, ..., 0, 0, 0],
                [0, 0, 0, ..., 0, 0, 0]])
```

```
In [20]: from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB()
model.fit(X_train_count, y_train)
```

```
Out[20]: MultinomialNB()
```

```
In [21]: emails = ['How are you brother?', 'Get Flat 50% on Your body outfits']
          emails_count = v.transform(emails)
          model.predict(emails_count)
```

```
Out[21]: array([0, 1])
```

```
In [22]: X_test_count = v.transform(X_test)
          model.score(X_test_count, y_test)
```

Out[22]: 0.990667623833453

In [ ]: