

```
import pandas as pd
```

```
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.metrics import ConfusionMatrixDisplay, accuracy_score, classification_report
```

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import CountVectorizer
```

```
df = pd.read_csv('/content/SMSSpamCollection', sep='\t', names=['label', 'message'])
```

```
df
```

	label	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 × 2 columns

```
df['label'] = df['label'].map({'ham': 0, 'spam': 1})
```

```
df
```

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

5572 rows × 2 columns

```
x = df['message'] #input
```

```
y = df['label'] #output
```

```
x
```

```

0      Go until jurong point, crazy.. Available only ...
1      Ok lar... Joking wif u oni...
2      Free entry in 2 a wkly comp to win FA Cup fina...
3      U dun say so early hor... U c already then say...
4      Nah I don't think he goes to usf, he lives aro...
...
5567   This is the 2nd time we have tried 2 contact u...
5568       Will ü b going to esplanade fr home?
5569   Pity, * was in mood for that. So...any other s...
5570   The guy did some bitching but I acted like i'd...
5571       Rofl. Its true to its name
Name: message, Length: 5572, dtype: object

```

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test = train_test_split(x,y,random_state=0,test_size=0.2)
```

```
x_train
```

```

1114   No I'm good for the movie, is it ok if I leave...
3589   If you were/are free i can give. Otherwise nal...
3095   Have you emigrated or something? Ok maybe 5.30...
1012       I just got home babe, are you still awake ?
3320       Kay... Since we are out already
...
4931   Hi, the SEXYCHAT girls are waiting for you to ...
3264       So u gonna get deus ex?
1653   For ur chance to win a £250 cash every wk TXT:...
2607   R U &SAM P IN EACHOTHER. IF WE MEET WE CAN GO ...
2732   Mm feeling sleepy. today itself i shall get th...
Name: message, Length: 4457, dtype: object

```

```
y_train
```

```

1114   0
3589   0
3095   0
1012   0
3320   0
..
4931   1
3264   0
1653   1
2607   0
2732   0
Name: label, Length: 4457, dtype: int64

```

```
x_test
```

```

4456   Storming msg: Wen u lift d phne, u say "HELLO"...
690     <Forwarded from 448712404000>Please CALL 08712...
944     And also I've sorta blown him off a couple tim...
3768       Sir Goodmorning, Once free call me.
1189   All will come alive.better correct any good lo...
...
2906   Ha. You don't know either. I did a a clever bu...
1270       Tee hee. Off to lecture, cheery bye bye.
3944   I got a call from a landline number. . . I am ...
2124   +123 Congratulations - in this week's competit...
253       What you doing?how are you?
Name: message, Length: 1115, dtype: object

```

```
vectorizer = CountVectorizer()
```

```

x_train_count = vectorizer.fit_transform(x_train)
x_test_count = vectorizer.transform(x_test)

```

```
model=MultinomialNB()
```

```
model.fit(x_train_count,y_train)
```

```

+ MultinomialNB
MultinomialNB()

```

```
y_pred=model.predict(x_test_count)

accuracy_score(y_test,y_pred)

0.9874439461883409

print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.99	1.00	0.99	955
1	0.98	0.93	0.96	160
accuracy			0.99	1115
macro avg	0.98	0.96	0.97	1115
weighted avg	0.99	0.99	0.99	1115

```
cross_val_score(model,x_train_count,y_train,cv=5)

array([0.98206278, 0.97309417, 0.97867565, 0.98316498, 0.98316498])
```

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.