

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.metrics import roc_auc_score, f1_score, confusion_matrix
from sklearn.naive_bayes import MultinomialNB
```

```
In [2]: data = pd.read_csv('spam.csv',encoding = 'ISO-8859-1')
data
```

Out[2]:

	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
5	spam	FreeMsg Hey there darling it's been 3 week's n...	NaN	NaN	NaN
6	ham	Even my brother is not like to speak with me. ...	NaN	NaN	NaN
7	ham	As per your request 'Melle Melle (Oru Minnamin...	NaN	NaN	NaN
8	spam	WINNER!! As a valued network customer you have...	NaN	NaN	NaN
9	spam	Had your mobile 11 months or more? U R entitle...	NaN	NaN	NaN
10	ham	I'm gonna be home soon and i don't want to tal...	NaN	NaN	NaN
11	spam	SIX chances to win CASH! From 100 to 20,000 po...	NaN	NaN	NaN
12	spam	URGENT! You have won a 1 week FREE membership ...	NaN	NaN	NaN
13	ham	I've been searching for the right words to tha...	NaN	NaN	NaN
14	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	NaN	NaN	NaN
15	spam	XXXMobileMovieClub: To use your credit, click ...	NaN	NaN	NaN
16	ham	Oh k...i'm watching here:)	NaN	NaN	NaN
17	ham	Eh u remember how 2 spell his name... Yes i di...	NaN	NaN	NaN
18	ham	Fine if that's the way u feel. That's the wa...	NaN	NaN	NaN
19	spam	England v Macedonia - dont miss the goals/team...	NaN	NaN	NaN
20	ham	Is that seriously how you spell his name?	NaN	NaN	NaN
21	ham	I'm going to try for 2 months ha ha only joking	NaN	NaN	NaN
22	ham	So I _ pay first lar... Then when is da stock c...	NaN	NaN	NaN
23	ham	Aft i finish my lunch then i go str down lor. ...	NaN	NaN	NaN
24	ham	Fffffff. Alright no way I can meet up with ...	NaN	NaN	NaN
25	ham	Just forced myself to eat a slice. I'm really ...	NaN	NaN	NaN
26	ham	Lol your always so convincing.	NaN	NaN	NaN
27	ham	Did you catch the bus ? Are you frying an egg ...	NaN	NaN	NaN
28	ham	I'm back & we're packing the car now, I'll...	NaN	NaN	NaN
29	ham	Ahhh. Work. I vaguely remember that! What does...	NaN	NaN	NaN
...
5542	ham	Armand says get your ass over to epsilon	NaN	NaN	NaN
5543	ham	U still havent got urself a jacket ah?	NaN	NaN	NaN
5544	ham	I'm taking derek & taylor to walmart, if I...	NaN	NaN	NaN
5545	ham	Hi its in durban are you still on this number	NaN	NaN	NaN
5546	ham	lc. There are a lotta childporn cars then.	NaN	NaN	NaN
5547	spam	Had your contract mobile 11 Mnths? Latest Moto...	NaN	NaN	NaN
5548	ham	No, I was trying it all weekend ;V	NaN	NaN	NaN
5549	ham	You know, wot people wear. T shirts, jumpers, ...	NaN	NaN	NaN

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
5550	ham	Cool, what time you think you can get here?	NaN	NaN	NaN
5551	ham	Wen did you get so spiritual and deep. That's ...	NaN	NaN	NaN
5552	ham	Have a safe trip to Nigeria. Wish you happines...	NaN	NaN	NaN
5553	ham	Hahaha..use your brain dear	NaN	NaN	NaN
5554	ham	Well keep in mind I've only got enough gas for...	NaN	NaN	NaN
5555	ham	Yeh. Indians was nice. Tho it did kane me off ...	NaN	NaN	NaN
5556	ham	Yes i have. So that's why u texted. Pshew...mi...	NaN	NaN	NaN
5557	ham	No. I meant the calculation is the same. That ...	NaN	NaN	NaN
5558	ham	Sorry, I'll call later	NaN	NaN	NaN
5559	ham	if you aren't here in the next &#x26;#x26;hou...	NaN	NaN	NaN
5560	ham	Anything lor. Juz both of us lor.	NaN	NaN	NaN
5561	ham	Get me out of this dump heap. My mom decided t...	NaN	NaN	NaN
5562	ham	Ok lor... Sony ericsson salesman... I ask shuh...	NaN	NaN	NaN
5563	ham	Ard 6 like dat lor.	NaN	NaN	NaN
5564	ham	Why don't you wait 'til at least wednesday to ...	NaN	NaN	NaN
5565	ham	Huh y lei...	NaN	NaN	NaN
5566	spam	REMINDER FROM O2: To get 2.50 pounds free call...	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 × 5 columns

In [3]: data.head()

Out[3]:

	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

In [4]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 5 columns):
v1          5572 non-null object
v2          5572 non-null object
Unnamed: 2   50 non-null object
Unnamed: 3   12 non-null object
Unnamed: 4    6 non-null object
dtypes: object(5)
memory usage: 217.7+ KB
```

In [5]: `data.describe`

Out[5]: <bound method NDFrame.describe of v1

v2	Unnamed: 2	\	
0	ham	Go until jurong point, crazy.. Available only ...	NaN
1	ham	Ok lar... Joking wif u oni...	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN
3	ham	U dun say so early hor... U c already then say...	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN
5	spam	FreeMsg Hey there darling it's been 3 week's n...	NaN
6	ham	Even my brother is not like to speak with me. ...	NaN
7	ham	As per your request 'Melle Melle (Oru Minnamin...	NaN
8	spam	WINNER!! As a valued network customer you have...	NaN
9	spam	Had your mobile 11 months or more? U R entitle...	NaN
10	ham	I'm gonna be home soon and i don't want to tal...	NaN
11	spam	SIX chances to win CASH! From 100 to 20,000 po...	NaN
12	spam	URGENT! You have won a 1 week FREE membership ...	NaN
13	ham	I've been searching for the right words to tha...	NaN
14	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	NaN
15	spam	XXXMobileMovieClub: To use your credit, click ...	NaN
16	ham	Oh k...i'm watching here:)	NaN
17	ham	Remember how 2 spell his name? Yes i do	NaN

In [6]: `data = data.drop(columns=data.columns[2:5])`
`data.head()`

Out[6]:

	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...

```
In [7]: data.columns = ['Category', 'Message']  
data
```

Out[7]:

	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...
5	spam	FreeMsg Hey there darling it's been 3 week's n...
6	ham	Even my brother is not like to speak with me. ...
7	ham	As per your request 'Melle Melle (Oru Minnamin...
8	spam	WINNER!! As a valued network customer you have...
9	spam	Had your mobile 11 months or more? U R entitle...
10	ham	I'm gonna be home soon and i don't want to tal...
11	spam	SIX chances to win CASH! From 100 to 20,000 po...
12	spam	URGENT! You have won a 1 week FREE membership ...
13	ham	I've been searching for the right words to tha...
14	ham	I HAVE A DATE ON SUNDAY WITH WILL!!
15	spam	XXXMobileMovieClub: To use your credit, click ...
16	ham	Oh k...i'm watching here:)
17	ham	Eh u remember how 2 spell his name... Yes i di...
18	ham	Fine if that's the way u feel. That's the wa...
19	spam	England v Macedonia - dont miss the goals/team...
20	ham	Is that seriously how you spell his name?
21	ham	I'm going to try for 2 months ha ha only joking
22	ham	So I _ pay first lar... Then when is da stock c...
23	ham	Aft i finish my lunch then i go str down lor. ...
24	ham	Fffffff. Alright no way I can meet up with ...
25	ham	Just forced myself to eat a slice. I'm really ...
26	ham	Lol your always so convincing.
27	ham	Did you catch the bus ? Are you frying an egg ...
28	ham	I'm back & we're packing the car now, I'll...
29	ham	Ahhh. Work. I vaguely remember that! What does...
...
5542	ham	Armand says get your ass over to epsilon
5543	ham	U still havent got urself a jacket ah?
5544	ham	I'm taking derek & taylor to walmart, if I...
5545	ham	Hi its in durban are you still on this number
5546	ham	lc. There are a lotta childporn cars then.
5547	spam	Had your contract mobile 11 Mnths? Latest Moto...
5548	ham	No, I was trying it all weekend ;V
5549	ham	You know, wot people wear. T shirts, jumpers, ...

	Category	Message
5550	ham	Cool, what time you think you can get here?
5551	ham	Wen did you get so spiritual and deep. That's ...
5552	ham	Have a safe trip to Nigeria. Wish you happines...
5553	ham	Hahaha..use your brain dear
5554	ham	Well keep in mind I've only got enough gas for...
5555	ham	Yeh. Indians was nice. Tho it did kane me off ...
5556	ham	Yes i have. So that's why u texted. Pshew...mi...
5557	ham	No. I meant the calculation is the same. That ...
5558	ham	Sorry, I'll call later
5559	ham	if you aren't here in the next & hou...
5560	ham	Anything lor. Juz both of us lor.
5561	ham	Get me out of this dump heap. My mom decided t...
5562	ham	Ok lor... Sony ericsson salesman... I ask shuh...
5563	ham	Ard 6 like dat lor.
5564	ham	Why don't you wait 'til at least wednesday to ...
5565	ham	Huh y lei...
5566	spam	REMINDER FROM O2: To get 2.50 pounds free call...
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

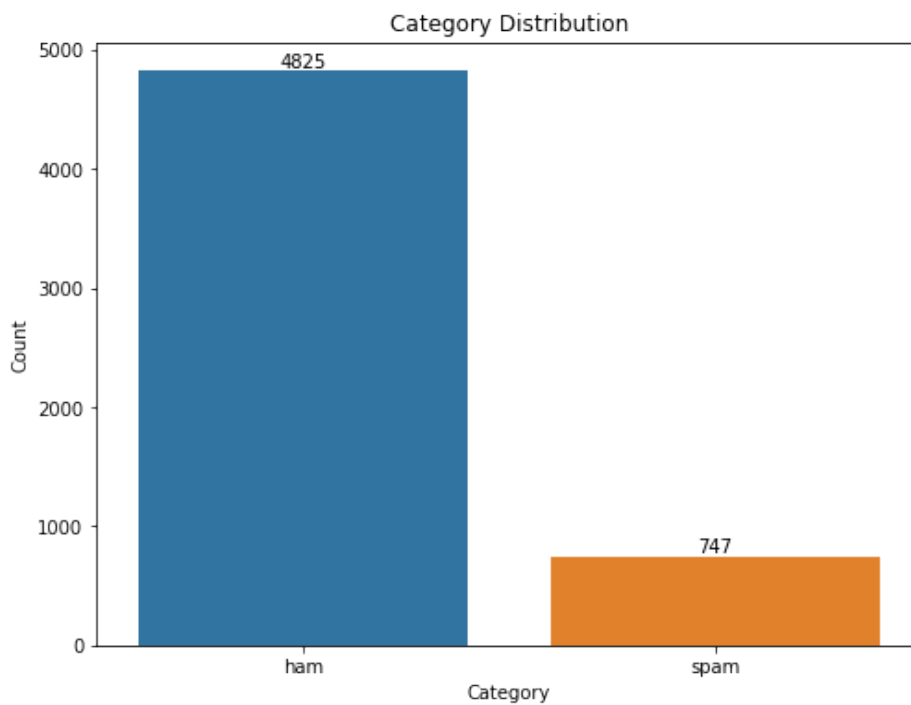
5572 rows × 2 columns

In [8]: `data.isnull().sum()`

Out[8]: Category 0
 Message 0
 dtype: int64


```
In [9]: category_counts = data['Category'].value_counts().reset_index()
category_counts.columns = ['Category', 'Count']
plt.figure(figsize=(8, 6))
sns.barplot(x='Category', y='Count', data=category_counts)
plt.xlabel('Category')
plt.ylabel('Count')
plt.title('Category Distribution')

for i, count in enumerate(category_counts['Count']):
    plt.text(i, count, str(count), ha='center', va='bottom')
plt.show()
```



```
In [10]: data['spam'] = data['Category'].apply(lambda x: 1 if x=='spam' else 0)  
data
```

Out[10]:

	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
5	spam	FreeMsg Hey there darling it's been 3 week's n...	1
6	ham	Even my brother is not like to speak with me. ...	0
7	ham	As per your request 'Melle Melle (Oru Minnamin...	0
8	spam	WINNER!! As a valued network customer you have...	1
9	spam	Had your mobile 11 months or more? U R entitle...	1
10	ham	I'm gonna be home soon and i don't want to tal...	0
11	spam	SIX chances to win CASH! From 100 to 20,000 po...	1
12	spam	URGENT! You have won a 1 week FREE membership ...	1
13	ham	I've been searching for the right words to tha...	0
14	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	0
15	spam	XXXMobileMovieClub: To use your credit, click ...	1
16	ham	Oh k...i'm watching here:)	0
17	ham	Eh u remember how 2 spell his name... Yes i di...	0
18	ham	Fine if that's the way u feel. That's the wa...	0
19	spam	England v Macedonia - dont miss the goals/team...	1
20	ham	Is that seriously how you spell his name?	0
21	ham	I'm going to try for 2 months ha ha only joking	0
22	ham	So I _ pay first lar... Then when is da stock c...	0
23	ham	Aft i finish my lunch then i go str down lor. ...	0
24	ham	Fffffff. Alright no way I can meet up with ...	0
25	ham	Just forced myself to eat a slice. I'm really ...	0
26	ham	Lol your always so convincing.	0
27	ham	Did you catch the bus ? Are you frying an egg ...	0
28	ham	I'm back & we're packing the car now, I'll...	0
29	ham	Ahhh. Work. I vaguely remember that! What does...	0
...
5542	ham	Armand says get your ass over to epsilon	0
5543	ham	U still havent got urself a jacket ah?	0
5544	ham	I'm taking derek & taylor to walmart, if I...	0
5545	ham	Hi its in durban are you still on this number	0
5546	ham	lc. There are a lotta childporn cars then.	0
5547	spam	Had your contract mobile 11 Mnths? Latest Moto...	1
5548	ham	No, I was trying it all weekend ;V	0
5549	ham	You know, wot people wear. T shirts, jumpers, ...	0

	Category	Message	spam
5550	ham	Cool, what time you think you can get here?	0
5551	ham	Wen did you get so spiritual and deep. That's ...	0
5552	ham	Have a safe trip to Nigeria. Wish you happines...	0
5553	ham	Hahaha..use your brain dear	0
5554	ham	Well keep in mind I've only got enough gas for...	0
5555	ham	Yeh. Indians was nice. Tho it did kane me off ...	0
5556	ham	Yes i have. So that's why u texted. Pshew...mi...	0
5557	ham	No. I meant the calculation is the same. That ...	0
5558	ham	Sorry, I'll call later	0
5559	ham	if you aren't here in the next & hou...	0
5560	ham	Anything lor. Juz both of us lor.	0
5561	ham	Get me out of this dump heap. My mom decided t...	0
5562	ham	Ok lor... Sony ericsson salesman... I ask shuh...	0
5563	ham	Ard 6 like dat lor.	0
5564	ham	Why don't you wait 'til at least wednesday to ...	0
5565	ham	Huh y lei...	0
5566	spam	REMINDER FROM O2: To get 2.50 pounds free call...	1
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 × 3 columns

```
In [11]: data['spam'] = data['Category'].apply(lambda x: 1 if x=='spam' else 0)
data
```

Out[11]:

	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
5	spam	FreeMsg Hey there darling it's been 3 week's n...	1
6	ham	Even my brother is not like to speak with me. ...	0
7	ham	As per your request 'Melle Melle (Oru Minnamin...	0
8	spam	WINNER!! As a valued network customer you have...	1
9	spam	Had your mobile 11 months or more? U R entitle...	1
10	ham	I'm gonna be home soon and i don't want to tal...	0
11	spam	SIX chances to win CASH! From 100 to 20,000 po...	1
12	spam	URGENT! You have won a 1 week FREE membership ...	1
13	ham	I've been searching for the right words to tha...	0
14	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	0
15	spam	XXXMobileMovieClub: To use your credit, click ...	1
16	ham	Oh k...i'm watching here:)	0
17	ham	Eh u remember how 2 spell his name... Yes i di...	0
18	ham	Fine if that's the way u feel. That's the wa...	0
19	spam	England v Macedonia - dont miss the goals/team...	1
20	ham	Is that seriously how you spell his name?	0
21	ham	I'm going to try for 2 months ha ha only joking	0
22	ham	So I _ pay first lar... Then when is da stock c...	0
23	ham	Aft i finish my lunch then i go str down lor. ...	0
24	ham	Fffffff. Alright no way I can meet up with ...	0
25	ham	Just forced myself to eat a slice. I'm really ...	0
26	ham	Lol your always so convincing.	0
27	ham	Did you catch the bus ? Are you frying an egg ...	0
28	ham	I'm back & we're packing the car now, I'll...	0
29	ham	Ahhh. Work. I vaguely remember that! What does...	0
...
5542	ham	Armand says get your ass over to epsilon	0
5543	ham	U still havent got urself a jacket ah?	0
5544	ham	I'm taking derek & taylor to walmart, if I...	0
5545	ham	Hi its in durban are you still on this number	0
5546	ham	lc. There are a lotta childporn cars then.	0
5547	spam	Had your contract mobile 11 Mnths? Latest Moto...	1
5548	ham	No, I was trying it all weekend ;V	0
5549	ham	You know, wot people wear. T shirts, jumpers, ...	0

	Category	Message	spam
5550	ham	Cool, what time you think you can get here?	0
5551	ham	Wen did you get so spiritual and deep. That's ...	0
5552	ham	Have a safe trip to Nigeria. Wish you happines...	0
5553	ham	Hahaha..use your brain dear	0
5554	ham	Well keep in mind I've only got enough gas for...	0
5555	ham	Yeh. Indians was nice. Tho it did kane me off ...	0
5556	ham	Yes i have. So that's why u texted. Pshew...mi...	0
5557	ham	No. I meant the calculation is the same. That ...	0
5558	ham	Sorry, I'll call later	0
5559	ham	if you aren't here in the next <#> hou...	0
5560	ham	Anything lor. Juz both of us lor.	0
5561	ham	Get me out of this dump heap. My mom decided t...	0
5562	ham	Ok lor... Sony ericsson salesman... I ask shuh...	0
5563	ham	Ard 6 like dat lor.	0
5564	ham	Why don't you wait 'til at least wednesday to ...	0
5565	ham	Huh y lei...	0
5566	spam	REMINDER FROM O2: To get 2.50 pounds free call...	1
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	Roff. Its true to its name	0

5572 rows × 3 columns

```
In [13]: X_train, X_test, y_train, y_test = train_test_split(data.Message,data.spam, test_size=0.2
from sklearn.feature_extraction.text import CountVectorizer
featurer = CountVectorizer()
X_train_count = featurer.fit_transform(X_train.values)
```

```
In [14]: X_train_count
```

```
Out[14]: <4457x7700 sparse matrix of type '<class 'numpy.int64'>'
with 59006 stored elements in Compressed Sparse Row format>
```

```
In [15]: model = MultinomialNB()
model.fit(X_train_count,y_train)
```

```
Out[15]: MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True)
```

```
In [16]: X_test_count = featurer.transform(X_test)
model.score(X_test_count, y_test)
```

```
Out[16]: 0.9838565022421525
```

```
In [17]: from sklearn.pipeline import Pipeline
         clf = Pipeline([
             ('vectorizer', CountVectorizer()),
             ('nb', MultinomialNB())
         ])
```

```
In [18]: clf.fit(X_train, y_train)
```

```
Out[18]: Pipeline(memory=None,
                 steps=[('vectorizer', CountVectorizer(analyzer='word', binary=False, decode_error='strict',
                 dtype=<class 'numpy.int64'>, encoding='utf-8', input='content',
                 lowercase=True, max_df=1.0, max_features=None, min_df=1,
                 ngram_range=(1, 1), preprocessor=None, stop_words=None,
                 strip_accents=None, token_pattern='(?u)\\b\\w\\w+\\b',
                 tokenizer=None, vocabulary=None)), ('nb', MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True))])
```

```
In [19]: clf.score(X_test,y_test)
```

```
Out[19]: 0.9838565022421525
```

```
In [20]: # Pre-trained model
         pretrained_model = model
         new_sentences = [
             "Your account have 100 debeted, is waiting to be collected. Simply text the password
         ]

         new_sentences_count = featurer.transform(new_sentences)
         # Predict whether each sentence is spam (1) or not (0)
         predictions = pretrained_model.predict(new_sentences_count)

         for sentence, prediction in zip(new_sentences, predictions):
             if prediction == 1:
                 print(f"'{sentence}' is a spam message.")
             else:
                 print(f"'{sentence}' is not a spam message.")
```

'Your account have 100 debeted, is waiting to be collected. Simply text the password \M IX" to 85069 to verify. Get Usher and Britney. FML' is a spam message.

```
In [ ]:
```