In [6]:
```
import pandas as pd
```

Load The Dataset

In [7]:
```
df=pd.read_csv(r"C:\Users\Biswajeet Jena\Documents\Csv\Spam Detection Datase
df
```

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...
...	...	...
5567	spam	This is the 2nd time we have tried 2 contact u...
5568	ham	Will Ì? 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]:
```
df.groupby('Category').describe()
```

Out[8]:

	Message			
	count	**unique**	**top**	**freq**
Category				
ham	4825	4516	Sorry, I'll call later	30
spam	747	641	Please call our customer service representativ...	4

Label the dataset

In [9]:
```
from sklearn.preprocessing import LabelEncoder
l=LabelEncoder()
l.fit(df.Category)
df['spam']=l.transform(df.Category)
df
```

Out[9]:

	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
...	...	...	...
5567	spam	This is the 2nd time we have tried 2 contact u...	1
5568	ham	Will Ì? 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

Split the dataset

In [10]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(df.Message,df.spam,test_size=
```

In [11]:

```
x_train
```

Out[11]:

```
2572    Û? and donÛ÷t worry weÛ÷ll have finished by...
2233    Nothing just getting msgs by dis name wit diff...
4583    Wow didn't think it was that common. I take it...
3666                Ha... U jus ate honey ar? So sweet...
5265                           Gud ni8.swt drms.take care
                                  ...
847     My stomach has been thru so much trauma I swea...
285     Yeah I think my usual guy's still passed out f...
2857    Japanese Proverb: If one Can do it, U too Can ...
1159                  Hey! There's veggie pizza... :/
483                                  Watching tv lor...
Name: Message, Length: 4457, dtype: object
```

In [12]:

```
y_train
```

Out[12]:

```
2572    0
2233    0
4583    0
3666    0
5265    0
       ..
847     0
285     0
2857    0
1159    0
483     0
Name: spam, Length: 4457, dtype: int32
```

In [13]:

```
x_test
```

Out[13]:

```
1170    Msgs r not time pass.They silently say that I ...
3195    And you! Will expect you whenever you text! Ho...
3411    Joy's father is John. Then John is the ____ of...
5395    Dunno lei shd b driving lor cos i go sch 1 hr ...
2506                 Congrats kano..whr s the treat maga?
                              ...
3651    We are hoping to get away by 7, from Langport....
4245    Aight, I'm chillin in a friend's room so text ...
5146    Oh unintentionally not bad timing. Great. Fing...
3862    Free Msg: Ringtone!From: http://tms. widelive....
2334                    What happen to her tell the truth
Name: Message, Length: 1115, dtype: object
```

In [14]:

```
y_test
```

Out[14]:

```
1170    0
3195    0
3411    0
5395    0
2506    0
       ..
3651    0
4245    0
5146    0
3862    1
2334    0
Name: spam, Length: 1115, dtype: int32
```

Convert the text data into matrix

In [15]:

```
from sklearn.feature_extraction.text import CountVectorizer
cv= CountVectorizer()
train_data=cv.fit_transform(x_train.values)
train_data
```

Out[15]:

```
<4457x7861 sparse matrix of type '<class 'numpy.int64'>'
        with 59333 stored elements in Compressed Sparse Row format>
```

In [16]:

```
test_data=cv.transform(x_test.values)
test_data
```

Out[16]:

```
<1115x7861 sparse matrix of type '<class 'numpy.int64'>'
	with 13879 stored elements in Compressed Sparse Row format>
```

Built a model using Naive Bayes and train it

In [17]:

```
from sklearn.naive_bayes import MultinomialNB
model=MultinomialNB()
model.fit(train_data,y_train)
```

Out[17]:

```
▼ MultinomialNB
MultinomialNB()
```

Check the accuracy of this model

In [18]:

```
model.score(test_data,y_test)
```

Out[18]:

```
0.9928251121076234
```

Now the model is ready to predict

In [19]:

```
model.predict(test_data[:10])
```

Out[19]:

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

In [20]:

```
y_test[:10]
```

Out[20]:

```
1170    0
3195    0
3411    0
5395    0
2506    0
5192    0
2634    0
1652    0
2325    0
2141    0
Name: spam, dtype: int32
```

Here we can see our model is too good. it predictd the first ten values of test perfectly

In [ ]:

```

```

Loading [MathJax]/extensions/Safe.js

Lets understand the model's performance using Confusion Matrix

In [21]:

```
from sklearn.metrics import confusion_matrix
```

In [22]:

```
predicted_value=model.predict(test_data)

cm=confusion_matrix(y_test,predicted_value)
cm
```

Out[22]:

```
array([[974,   1],
       [  7, 133]], dtype=int64)
```

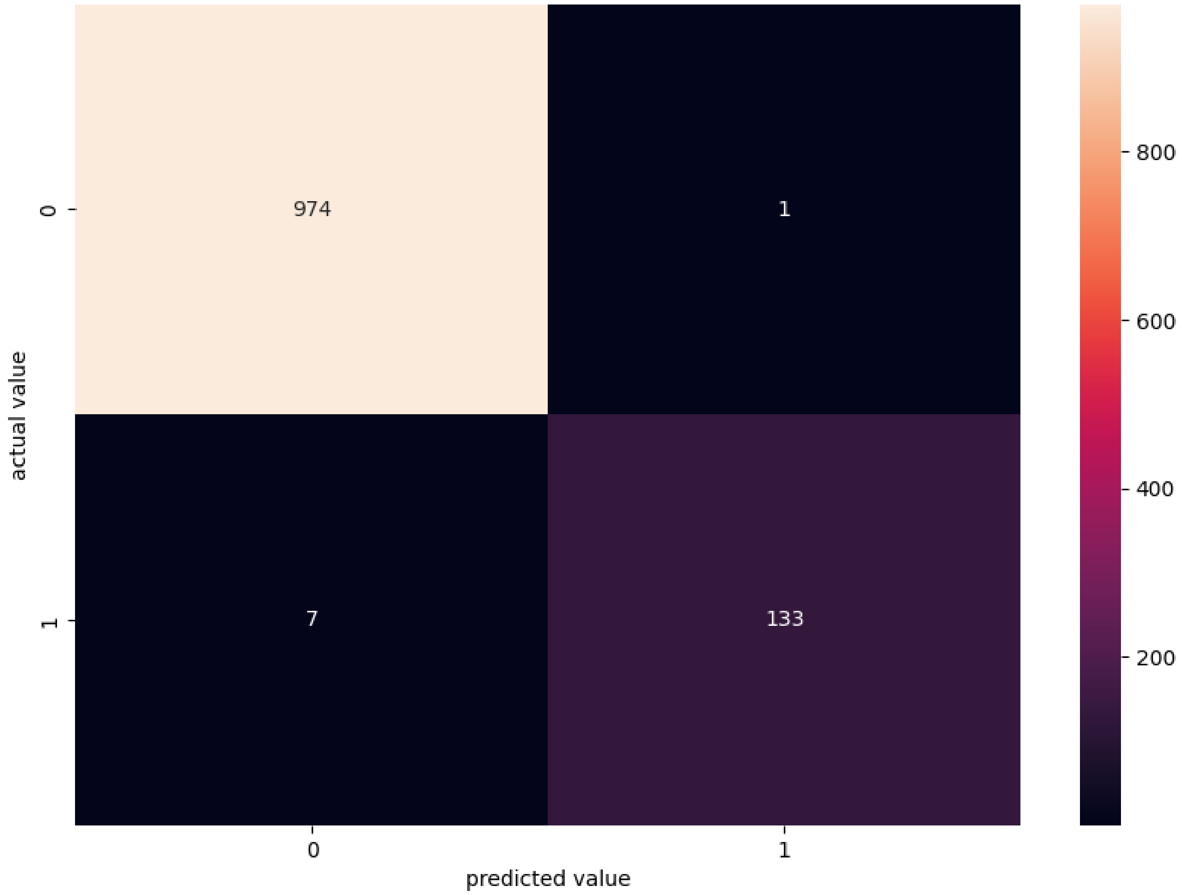
Lets understand it Visually

In [23]:

```
import matplotlib.pyplot as plt
import seaborn as sb
plt.figure(figsize=(10,7))
sb.heatmap(cm,annot=True,fmt='d')
plt.xlabel('predicted value')
plt.ylabel('actual value')
```

Out[23]:

```
Text(95.72222222222221, 0.5, 'actual value')
```



Here we can see 981 times it predict 0 and for 974 times it predict right and for 7 times it predict wrong. Simillarly 134 times it predict 1 and for 133 times it predict right and for 1 times it predict wrong

```
In [ ]:
```

```
In [ ]:
```