In [1]:

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

```

```
/kaggle/input/spam-email/spam.csv
```

In [2]:

```
data=pd.read_csv('spam.csv')
data
```

Out[2]:

	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 × 2 columns

In [3]:

```
data.columns
```

Out[3]:

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

In [4]:

```
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
```

Dropped The Column Unnamed: 0

In [5]:

```
data.isna().sum()
```

Out[5]:

```
Category    0
Message     0
dtype: int64
```

In [6]:

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

Out[6]:

	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

In [7]:

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(data.Message,data.Spam,t
```

In [8]:

```
#CounterVectorizer Convert the text into matrics
from sklearn.feature_extraction.text import CountVectorizer
```

Naive Bayes Have three Classifier(Bernouli,Multinominal,Gaussian) Here I use Multinominal Bayes Because here data in a discrete form discrete data(e.g movie ratings ranging 1 to 5 as each rating will have certain frequency to represent)

In [9]:

```
from sklearn.naive_bayes import MultinomialNB
```

In [10]:

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

Tarining The Model

In [11]:

```
clf.fit(X_train,y_train)
```

Out[11]:

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

Here I given Two email Two detect 1st One is looking good and the other one looking spam

In [12]:

```
emails=[
    'Sounds great! Are you home now?',
    'Will u meet ur dream partner soon? Is ur career off 2 a flyng star
]
```

Predict Email

In [13]:

```
clf.predict(emails)
```

Out[13]:

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

Prediction Of Model

In [14]:

```
clf.score(X_test,y_test)
```

Out[14]:

```
0.9777458722182341
```