In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
data =   {
    'text': [
        'NLP is not a interesting subject',
        'I love programming in Python',
        'Python is great for machine learning',
        'I love learning new things',
        'Data science is my favorite subject',
        'Machine learning models are fascinating',
        'I am excited about AI advancements',
        'Building models is very rewarding',
        'Deep learning is very interesting',
        'Exploring new algorithms is fun',
        'I like solving complex problems',
        'Python is a powerful language',
        'Working with data is exciting',
        'Creating models is challenging',
        'I love analyzing data patterns',
        'Predictive modeling is fascinating',
        'I hate bugs in code',
        'Debugging code can be frustrating',
        'Errors in code are annoying',
        'Fixing bugs gives me a sense of achievement',
        'I dislike syntax errors',
        'Data cleaning is tedious',
        'Working with dirty data is challenging',
        'Model performance issues are frustrating',
        'I find error messages confusing',
        'I get frustrated with slow code',
        'Handling missing data is tough',
        'Training models can be exhausting',
        'Dealing with noisy data is hard',
        'Code optimization is difficult',
        'I dislike inconsistent data formats',"Data cleaning is happy"
    ],
    'label': [
        'negative','positive', 'positive', 'positive', 'positive', 'positive',
        'positive', 'positive', 'positive', 'positive', 'positive',
        'positive', 'positive', 'positive', 'positive', 'positive',
        'negative', 'negative', 'negative', 'negative', 'negative',
        'negative', 'negative', 'negative', 'negative', 'negative',
        'negative', 'negative', 'negative', 'negative', 'negative',"positive"
    ]
}
# Load data into a DataFrame
df = pd.DataFrame(data)
df
```

Out[2]:

	text	label
0	NLP is not a interesting subject	negative
1	I love programming in Python	positive
2	Python is great for machine learning	positive
3	I love learning new things	positive
4	Data science is my favorite subject	positive
5	Machine learning models are fascinating	positive
6	I am excited about AI advancements	positive
7	Building models is very rewarding	positive
8	Deep learning is very interesting	positive
9	Exploring new algorithms is fun	positive
10	I like solving complex problems	positive
11	Python is a powerful language	positive
12	Working with data is exciting	positive
13	Creating models is challenging	positive
14	I love analyzing data patterns	positive
15	Predictive modeling is fascinating	positive
16	I hate bugs in code	negative
17	Debugging code can be frustrating	negative
18	Errors in code are annoying	negative
19	Fixing bugs gives me a sense of achievement	negative
20	I dislike syntax errors	negative
21	Data cleaning is tedious	negative
22	Working with dirty data is challenging	negative
23	Model performance issues are frustrating	negative
24	I find error messages confusing	negative
25	I get frustrated with slow code	negative
26	Handling missing data is tough	negative
27	Training models can be exhausting	negative
28	Dealing with noisy data is hard	negative
29	Code optimization is difficult	negative
30	I dislike inconsistent data formats	negative
31	Data cleaning is happy	positive

In [3]:
```
x=df.text
y=df.label
x
```

Out[3]:
```
0                NLP is not a interesting subject
1                    I love programming in Python
2            Python is great for machine learning
3                      I love learning new things
4             Data science is my favorite subject
5         Machine learning models are fascinating
6              I am excited about AI advancements
7               Building models is very rewarding
8               Deep learning is very interesting
9                 Exploring new algorithms is fun
10                I like solving complex problems
11                  Python is a powerful language
12                  Working with data is exciting
13                Creating models is challenging
14                I love analyzing data patterns
15            Predictive modeling is fascinating
16                           I hate bugs in code
17              Debugging code can be frustrating
18                    Errors in code are annoying
19    Fixing bugs gives me a sense of achievement
20                        I dislike syntax errors
21                       Data cleaning is tedious
22         Working with dirty data is challenging
23       Model performance issues are frustrating
24                I find error messages confusing
25                I get frustrated with slow code
26                 Handling missing data is tough
27              Training models can be exhausting
28                Dealing with noisy data is hard
29                 Code optimization is difficult
30           I dislike inconsistent data formats
31                         Data cleaning is happy
Name: text, dtype: object
```

In [4]:
```
from sklearn.feature_extraction.text import  CountVectorizer
cv=CountVectorizer()
x1=cv.fit_transform(x)
```

In [5]:
```
x1
```

Out[5]:
```
<32x92 sparse matrix of type '<class 'numpy.int64'>'
        with 151 stored elements in Compressed Sparse Row format>
```

In [6]:
```
x1.toarray()
```

Out[6]:
```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

In [7]:
```
cv.get_feature_names_out()
```

Out[7]:
```
array(['about', 'achievement', 'advancements', 'ai', 'algorithms', 'am',
       'analyzing', 'annoying', 'are', 'be', 'bugs', 'building', 'can',
       'challenging', 'cleaning', 'code', 'complex', 'confusing',
       'creating', 'data', 'dealing', 'debugging', 'deep', 'difficult',
       'dirty', 'dislike', 'error', 'errors', 'excited', 'exciting',
       'exhausting', 'exploring', 'fascinating', 'favorite', 'find',
       'fixing', 'for', 'formats', 'frustrated', 'frustrating', 'fun',
       'get', 'gives', 'great', 'handling', 'happy', 'hard', 'hate', 'in',
       'inconsistent', 'interesting', 'is', 'issues', 'language',
       'learning', 'like', 'love', 'machine', 'me', 'messages', 'missing',
       'model', 'modeling', 'models', 'my', 'new', 'nlp', 'noisy', 'not',
       'of', 'optimization', 'patterns', 'performance', 'powerful',
       'predictive', 'problems', 'programming', 'python', 'rewarding',
       'science', 'sense', 'slow', 'solving', 'subject', 'syntax',
       'tedious', 'things', 'tough', 'training', 'very', 'with',
       'working'], dtype=object)
```

In [8]:
```
pd.DataFrame(x1.toarray(),columns=cv.get_feature_names_out())
```

Out[8]:

	about	achievement	advancements	ai	algorithms	am	analyzing	annoying	are	be	...	solving	subject	syntax	tedious	things	to
0	0	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	
4	0	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0	
5	0	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	
6	1	0	1	1	0	1	0	0	0	0	...	0	0	0	0	0	
7	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
9	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0	
10	0	0	0	0	0	0	0	0	0	0	...	1	0	0	0	0	
11	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
12	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
13	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
14	0	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
16	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
17	0	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0	
18	0	0	0	0	0	0	0	1	1	0	...	0	0	0	0	0	
19	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
20	0	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	
21	0	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	
22	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
23	0	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	
24	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
25	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
26	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
27	0	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0	
28	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
29	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
30	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
31	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	

32 rows × 92 columns

In [9]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x1,y,test_size=0.2)
```

In [10]:
```
x_train
```

Out[10]:
```
<25x92 sparse matrix of type '<class 'numpy.int64'>'
        with 115 stored elements in Compressed Sparse Row format>
```

In [11]:
```
x_test
```

Out[11]:
```
<7x92 sparse matrix of type '<class 'numpy.int64'>'
        with 36 stored elements in Compressed Sparse Row format>
```

In [12]:
```
x_train.shape
```

Out[12]:
```
(25, 92)
```

In [13]:
```
x_test.shape
```

Out[13]:
```
(7, 92)
```

In [14]:
```
#SUPERVISED LEARNING:CLASSIFICATION:LOGISTIC REGRESSION,RANDOM FOREST CLASSIFIER,SVC,KNN,DECISION TREE CLASSIFIEI
```

In [15]:
```
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression()
lr.fit(x_train,y_train)
```

Out[15]:
```
LogisticRegression
LogisticRegression()
```

In [16]:
```
y_pred=lr.predict(x_test)
```

In [17]:
```
y_pred
```

Out[17]:
```
array(['negative', 'positive', 'positive', 'positive', 'negative',
       'negative', 'positive'], dtype=object)
```

In [18]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Out[18]:
```
0.7142857142857143
```

In [19]:
```
Z= ["NLP is a very interesting subject"]
```

In [20]:
```
U=cv.transform(Z)
U
```

Out[20]:
```
<1x92 sparse matrix of type '<class 'numpy.int64'>'
        with 5 stored elements in Compressed Sparse Row format>
```

In [21]:
```
lr.predict(U)
```

Out[21]:
```
array(['positive'], dtype=object)
```

In [22]:
```
N=["NLP is not a interesting subject"]
r=cv.transform(N)
lr.predict(r)
```

Out[22]:
```
array(['negative'], dtype=object)
```

In [ ]:
```

```