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

In [2]:
```python
True_news = pd.read_csv("true_news.csv")
Fake_news = pd.read_csv("false_news.csv")
```

In [3]:
```python
True_news
```

Out[3]:

	Title	Text	Subject	Date
0	True News Title 1	This is the true news text for entry 1, provid...	Health	2021-05-18
1	True News Title 2	This is the true news text for entry 2, provid...	Health	2020-05-15
2	True News Title 3	This is the true news text for entry 3, provid...	Sports	2022-07-11
3	True News Title 4	This is the true news text for entry 4, provid...	Science	2023-04-15
4	True News Title 5	This is the true news text for entry 5, provid...	Politics	2023-06-16
5	True News Title 6	This is the true news text for entry 6, provid...	Politics	2020-06-25
6	True News Title 7	This is the true news text for entry 7, provid...	Entertainment	2023-04-23
7	True News Title 8	This is the true news text for entry 8, provid...	Entertainment	2021-12-01
8	True News Title 9	This is the true news text for entry 9, provid...	Politics	2020-09-28
9	True News Title 10	This is the true news text for entry 10, provi...	Sports	2021-12-26
10	True News Title 11	This is the true news text for entry 11, provi...	Politics	2021-06-28
11	True News Title 12	This is the true news text for entry 12, provi...	Health	2023-04-25
12	True News Title 13	This is the true news text for entry 13, provi...	Politics	2020-12-01
13	True News Title 14	This is the true news text for entry 14, provi...	Health	2020-06-22
14	True News Title 15	This is the true news text for entry 15, provi...	Entertainment	2020-04-27
15	True News Title 16	This is the true news text for entry 16, provi...	Technology	2021-11-21
16	True News Title 17	This is the true news text for entry 17, provi...	Politics	2023-08-19
17	True News Title 18	This is the true news text for entry 18, provi...	Politics	2022-09-20
18	True News Title 19	This is the true news text for entry 19, provi...	Technology	2023-12-17
19	True News Title 20	This is the true news text for entry 20, provi...	Technology	2023-06-26
20	True News Title 21	This is the true news text for entry 21, provi...	Science	2023-03-18
21	True News Title 22	This is the true news text for entry 22, provi...	Politics	2023-11-29
22	True News Title 23	This is the true news text for entry 23, provi...	Technology	2021-04-29
23	True News Title 24	This is the true news text for entry 24, provi...	Politics	2021-03-11
24	True News Title 25	This is the true news text for entry 25, provi...	Technology	2021-07-27
25	True News Title 26	This is the true news text for entry 26, provi...	Entertainment	2020-03-26
26	True News Title 27	This is the true news text for entry 27, provi...	Science	2023-03-10
27	True News Title 28	This is the true news text for entry 28, provi...	Health	2020-05-31
28	True News Title 29	This is the true news text for entry 29, provi...	Health	2020-06-08
29	True News Title 30	This is the true news text for entry 30, provi...	Sports	2023-02-09

In [4]: `Fake_news`

Out[4]:

	Title	Text	Subject	Date
0	False News Title 1	This is the false news text for entry 1, provi...	Technology	2023-08-02
1	False News Title 2	This is the false news text for entry 2, provi...	Science	2022-12-16
2	False News Title 3	This is the false news text for entry 3, provi...	Technology	2021-07-24
3	False News Title 4	This is the false news text for entry 4, provi...	Politics	2023-12-12
4	False News Title 5	This is the false news text for entry 5, provi...	Politics	2023-11-27
5	False News Title 6	This is the false news text for entry 6, provi...	Sports	2021-12-31
6	False News Title 7	This is the false news text for entry 7, provi...	Science	2022-04-23
7	False News Title 8	This is the false news text for entry 8, provi...	Science	2021-04-07
8	False News Title 9	This is the false news text for entry 9, provi...	Sports	2022-02-20
9	False News Title 10	This is the false news text for entry 10, prov...	Technology	2021-01-12
10	False News Title 11	This is the false news text for entry 11, prov...	Sports	2021-05-19
11	False News Title 12	This is the false news text for entry 12, prov...	Technology	2023-05-03
12	False News Title 13	This is the false news text for entry 13, prov...	Technology	2022-04-24
13	False News Title 14	This is the false news text for entry 14, prov...	Politics	2020-08-05
14	False News Title 15	This is the false news text for entry 15, prov...	Technology	2023-03-02
15	False News Title 16	This is the false news text for entry 16, prov...	Science	2022-03-14
16	False News Title 17	This is the false news text for entry 17, prov...	Technology	2020-02-19
17	False News Title 18	This is the false news text for entry 18, prov...	Sports	2022-07-15
18	False News Title 19	This is the false news text for entry 19, prov...	Technology	2022-05-02
19	False News Title 20	This is the false news text for entry 20, prov...	Technology	2020-06-30
20	False News Title 21	This is the false news text for entry 21, prov...	Health	2020-12-17
21	False News Title 22	This is the false news text for entry 22, prov...	Technology	2022-05-21
22	False News Title 23	This is the false news text for entry 23, prov...	Entertainment	2023-06-04
23	False News Title 24	This is the false news text for entry 24, prov...	Entertainment	2022-02-02
24	False News Title 25	This is the false news text for entry 25, prov...	Technology	2022-04-27
25	False News Title 26	This is the false news text for entry 26, prov...	Health	2023-10-28
26	False News Title 27	This is the false news text for entry 27, prov...	Politics	2023-03-20
27	False News Title 28	This is the false news text for entry 28, prov...	Science	2021-09-03
28	False News Title 29	This is the false news text for entry 29, prov...	Sports	2021-07-22
29	False News Title 30	This is the false news text for entry 30, prov...	Entertainment	2021-06-28

In [5]:
```python
True_news['label'] = 0
Fake_news['label'] = 1
```

In [6]:
```python
dataset1 = True_news[['Text','label']]
dataset2 = Fake_news[['Text','label']]
```

In [7]: `dataset1`

Out[7]:

	Text	label
0	This is the true news text for entry 1, provid...	0
1	This is the true news text for entry 2, provid...	0
2	This is the true news text for entry 3, provid...	0
3	This is the true news text for entry 4, provid...	0
4	This is the true news text for entry 5, provid...	0
5	This is the true news text for entry 6, provid...	0
6	This is the true news text for entry 7, provid...	0
7	This is the true news text for entry 8, provid...	0
8	This is the true news text for entry 9, provid...	0
9	This is the true news text for entry 10, provi...	0
10	This is the true news text for entry 11, provi...	0
11	This is the true news text for entry 12, provi...	0
12	This is the true news text for entry 13, provi...	0
13	This is the true news text for entry 14, provi...	0
14	This is the true news text for entry 15, provi...	0
15	This is the true news text for entry 16, provi...	0
16	This is the true news text for entry 17, provi...	0
17	This is the true news text for entry 18, provi...	0
18	This is the true news text for entry 19, provi...	0
19	This is the true news text for entry 20, provi...	0
20	This is the true news text for entry 21, provi...	0
21	This is the true news text for entry 22, provi...	0
22	This is the true news text for entry 23, provi...	0
23	This is the true news text for entry 24, provi...	0
24	This is the true news text for entry 25, provi...	0
25	This is the true news text for entry 26, provi...	0
26	This is the true news text for entry 27, provi...	0
27	This is the true news text for entry 28, provi...	0
28	This is the true news text for entry 29, provi...	0
29	This is the true news text for entry 30, provi...	0

In [8]: `dataset2`

Out[8]:

	Text	label
0	This is the false news text for entry 1, provi...	1
1	This is the false news text for entry 2, provi...	1
2	This is the false news text for entry 3, provi...	1
3	This is the false news text for entry 4, provi...	1
4	This is the false news text for entry 5, provi...	1
5	This is the false news text for entry 6, provi...	1
6	This is the false news text for entry 7, provi...	1
7	This is the false news text for entry 8, provi...	1
8	This is the false news text for entry 9, provi...	1
9	This is the false news text for entry 10, prov...	1
10	This is the false news text for entry 11, prov...	1
11	This is the false news text for entry 12, prov...	1
12	This is the false news text for entry 13, prov...	1
13	This is the false news text for entry 14, prov...	1
14	This is the false news text for entry 15, prov...	1
15	This is the false news text for entry 16, prov...	1
16	This is the false news text for entry 17, prov...	1
17	This is the false news text for entry 18, prov...	1
18	This is the false news text for entry 19, prov...	1
19	This is the false news text for entry 20, prov...	1
20	This is the false news text for entry 21, prov...	1
21	This is the false news text for entry 22, prov...	1
22	This is the false news text for entry 23, prov...	1
23	This is the false news text for entry 24, prov...	1
24	This is the false news text for entry 25, prov...	1
25	This is the false news text for entry 26, prov...	1
26	This is the false news text for entry 27, prov...	1
27	This is the false news text for entry 28, prov...	1
28	This is the false news text for entry 29, prov...	1
29	This is the false news text for entry 30, prov...	1

In [9]:
```python
dataset = pd.concat([dataset1,dataset2])
```

In [10]: dataset

Out[10]:

	Text	label
0	This is the true news text for entry 1, provid...	0
1	This is the true news text for entry 2, provid...	0
2	This is the true news text for entry 3, provid...	0
3	This is the true news text for entry 4, provid...	0
4	This is the true news text for entry 5, provid...	0
5	This is the true news text for entry 6, provid...	0
6	This is the true news text for entry 7, provid...	0
7	This is the true news text for entry 8, provid...	0
8	This is the true news text for entry 9, provid...	0
9	This is the true news text for entry 10, provi...	0
10	This is the true news text for entry 11, provi...	0
11	This is the true news text for entry 12, provi...	0
12	This is the true news text for entry 13, provi...	0
13	This is the true news text for entry 14, provi...	0
14	This is the true news text for entry 15, provi...	0
15	This is the true news text for entry 16, provi...	0
16	This is the true news text for entry 17, provi...	0
17	This is the true news text for entry 18, provi...	0
18	This is the true news text for entry 19, provi...	0
19	This is the true news text for entry 20, provi...	0
20	This is the true news text for entry 21, provi...	0
21	This is the true news text for entry 22, provi...	0
22	This is the true news text for entry 23, provi...	0
23	This is the true news text for entry 24, provi...	0
24	This is the true news text for entry 25, provi...	0
25	This is the true news text for entry 26, provi...	0
26	This is the true news text for entry 27, provi...	0
27	This is the true news text for entry 28, provi...	0
28	This is the true news text for entry 29, provi...	0
29	This is the true news text for entry 30, provi...	0
0	This is the false news text for entry 1, provi...	1
1	This is the false news text for entry 2, provi...	1
2	This is the false news text for entry 3, provi...	1
3	This is the false news text for entry 4, provi...	1
4	This is the false news text for entry 5, provi...	1
5	This is the false news text for entry 6, provi...	1
6	This is the false news text for entry 7, provi...	1
7	This is the false news text for entry 8, provi...	1
8	This is the false news text for entry 9, provi...	1
9	This is the false news text for entry 10, prov...	1
10	This is the false news text for entry 11, prov...	1
11	This is the false news text for entry 12, prov...	1

	Text	label
12	This is the false news text for entry 13, prov...	1
13	This is the false news text for entry 14, prov...	1
14	This is the false news text for entry 15, prov...	1
15	This is the false news text for entry 16, prov...	1
16	This is the false news text for entry 17, prov...	1
17	This is the false news text for entry 18, prov...	1
18	This is the false news text for entry 19, prov...	1
19	This is the false news text for entry 20, prov...	1
20	This is the false news text for entry 21, prov...	1
21	This is the false news text for entry 22, prov...	1
22	This is the false news text for entry 23, prov...	1
23	This is the false news text for entry 24, prov...	1
24	This is the false news text for entry 25, prov...	1
25	This is the false news text for entry 26, prov...	1
26	This is the false news text for entry 27, prov...	1
27	This is the false news text for entry 28, prov...	1
28	This is the false news text for entry 29, prov...	1
29	This is the false news text for entry 30, prov...	1

```
In [11]: dataset.shape
```

Out[11]: (60, 2)

```
In [12]: dataset.isnull().sum()
```

Out[12]: Text   0
         label   0
         dtype: int64

```
In [13]: dataset['label'].value_counts()
```

Out[13]: 1   30
         0   30
         Name: label, dtype: int64

```
In [14]: dataset['label'].value_counts()
```

Out[14]: 1   30
         0   30
         Name: label, dtype: int64

```
In [15]: dataset = dataset.sample(frac=1)
         dataset
```

Out[15]:

	Text	label
15	This is the true news text for entry 16, provi...	0
10	This is the false news text for entry 11, prov...	1
2	This is the true news text for entry 3, provid...	0
24	This is the false news text for entry 25, prov...	1
4	This is the true news text for entry 5, provid...	0
9	This is the true news text for entry 10, provi...	0
9	This is the false news text for entry 10, prov...	1
22	This is the false news text for entry 23, prov...	1
26	This is the true news text for entry 27, provi...	0
0	This is the true news text for entry 1, provid...	0
14	This is the false news text for entry 15, prov...	1
29	This is the true news text for entry 30, provi...	0
29	This is the false news text for entry 30, prov...	1
12	This is the false news text for entry 13, prov...	1
7	This is the true news text for entry 8, provid...	0
5	This is the false news text for entry 6, provi...	1
27	This is the true news text for entry 28, provi...	0
13	This is the true news text for entry 14, provi...	0
10	This is the true news text for entry 11, provi...	0
17	This is the true news text for entry 18, provi...	0
1	This is the false news text for entry 2, provi...	1
28	This is the true news text for entry 29, provi...	0
6	This is the false news text for entry 7, provi...	1
25	This is the false news text for entry 26, prov...	1
24	This is the true news text for entry 25, provi...	0
25	This is the true news text for entry 26, provi...	0
3	This is the false news text for entry 4, provi...	1
19	This is the true news text for entry 20, provi...	0
18	This is the false news text for entry 19, prov...	1
2	This is the false news text for entry 3, provi...	1
8	This is the false news text for entry 9, provi...	1
0	This is the false news text for entry 1, provi...	1
16	This is the true news text for entry 17, provi...	0
3	This is the true news text for entry 4, provid...	0
17	This is the false news text for entry 18, prov...	1
1	This is the true news text for entry 2, provid...	0
11	This is the true news text for entry 12, provi...	0
20	This is the false news text for entry 21, prov...	1
20	This is the true news text for entry 21, provi...	0
21	This is the false news text for entry 22, prov...	1
13	This is the false news text for entry 14, prov...	1
7	This is the false news text for entry 8, provi...	1

	Text	label
15	This is the false news text for entry 16, prov...	1
12	This is the true news text for entry 13, provi...	0
16	This is the false news text for entry 17, prov...	1
27	This is the false news text for entry 28, prov...	1
14	This is the true news text for entry 15, provi...	0
8	This is the true news text for entry 9, provid...	0
19	This is the false news text for entry 20, prov...	1
26	This is the false news text for entry 27, prov...	1
21	This is the true news text for entry 22, provi...	0
18	This is the true news text for entry 19, provi...	0
4	This is the false news text for entry 5, provi...	1
22	This is the true news text for entry 23, provi...	0
28	This is the false news text for entry 29, prov...	1
5	This is the true news text for entry 6, provid...	0
23	This is the true news text for entry 24, provi...	0
23	This is the false news text for entry 24, prov...	1
11	This is the false news text for entry 12, prov...	1
6	This is the true news text for entry 7, provid...	0

```python
In [16]:  import nltk
          import re
```

```python
In [17]:  from nltk.corpus import stopwords
```

```python
In [18]:  from nltk.stem import WordNetLemmatizer
```

```python
In [19]:  ps=WordNetLemmatizer
```

```python
In [20]:  stop_words = stopwords.words('english')
          stoplist = set(stopwords.words("english"))
```

```python
In [21]:  nltk.download('wordnet')
```

```
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\noora\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

Out[21]: True

```python
In [22]:  nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\noora\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[22]: True

In [23]:
```python
def clean_row(row):
    row = row.lower()
    row = re.sub('[^a-zA-Z]','',row)
    token = row.split()
    news = [ps.lemmatize(words) for words in token if not words in stopwords]
    cleanned_news = ''.join(news)
    return cleanned_news
```

In [32]:
```python
import re
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
import nltk

# Download necessary NLTK data
nltk.download('stopwords')
nltk.download('wordnet')

# Initialize lemmatizer and stopwords
ps = WordNetLemmatizer()
stopwords_list = set(stopwords.words('english'))

def clean_row(row):
    row = re.sub('[^a-zA-Z]', ' ', row)  # Remove non-alphabetic characters
    token = row.split()  # Tokenize the sentence
    news = [ps.lemmatize(word) for word in token if word.lower() not in stopwords_list
    cleaned_news = ' '.join(news)  # Recombine the words
    return cleaned_news

# Assuming 'dataset' is a pandas DataFrame
dataset['Text'] = dataset['Text'].apply(lambda x: clean_row(x))
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\noora\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\noora\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

In [34]: `dataset['Text']`

Out[34]:
```
15    true news text entry providing detailed descri...
10    false news text entry providing detailed descr...
2     true news text entry providing detailed descri...
24    false news text entry providing detailed descr...
4     true news text entry providing detailed descri...
9     true news text entry providing detailed descri...
9     false news text entry providing detailed descr...
22    false news text entry providing detailed descr...
26    true news text entry providing detailed descri...
0     true news text entry providing detailed descri...
14    false news text entry providing detailed descr...
29    true news text entry providing detailed descri...
29    false news text entry providing detailed descr...
12    false news text entry providing detailed descr...
7     true news text entry providing detailed descri...
5     false news text entry providing detailed descr...
27    true news text entry providing detailed descri...
13    true news text entry providing detailed descri...
10    true news text entry providing detailed descri...
17    true news text entry providing detailed descri...
1     false news text entry providing detailed descr...
28    true news text entry providing detailed descri...
6     false news text entry providing detailed descr...
25    false news text entry providing detailed descr...
24    true news text entry providing detailed descri...
25    true news text entry providing detailed descri...
3     false news text entry providing detailed descr...
19    true news text entry providing detailed descri...
18    false news text entry providing detailed descr...
2     false news text entry providing detailed descr...
8     false news text entry providing detailed descr...
0     false news text entry providing detailed descr...
16    true news text entry providing detailed descri...
3     true news text entry providing detailed descri...
17    false news text entry providing detailed descr...
1     true news text entry providing detailed descri...
11    true news text entry providing detailed descri...
20    false news text entry providing detailed descr...
20    true news text entry providing detailed descri...
21    false news text entry providing detailed descr...
13    false news text entry providing detailed descr...
7     false news text entry providing detailed descr...
15    false news text entry providing detailed descr...
12    true news text entry providing detailed descri...
16    false news text entry providing detailed descr...
27    false news text entry providing detailed descr...
14    true news text entry providing detailed descri...
8     true news text entry providing detailed descri...
19    false news text entry providing detailed descr...
26    false news text entry providing detailed descr...
21    true news text entry providing detailed descri...
18    true news text entry providing detailed descri...
4     false news text entry providing detailed descr...
22    true news text entry providing detailed descri...
28    false news text entry providing detailed descr...
5     true news text entry providing detailed descri...
23    true news text entry providing detailed descri...
23    false news text entry providing detailed descr...
11    false news text entry providing detailed descr...
6     true news text entry providing detailed descri...
Name: Text, dtype: object
```

```
In [35]: dataset['Text']
```

```
Out[35]: 15    true news text entry providing detailed descri...
         10    false news text entry providing detailed descr...
         2     true news text entry providing detailed descri...
         24    false news text entry providing detailed descr...
         4     true news text entry providing detailed descri...
         9     true news text entry providing detailed descri...
         9     false news text entry providing detailed descr...
         22    false news text entry providing detailed descr...
         26    true news text entry providing detailed descri...
         0     true news text entry providing detailed descri...
         14    false news text entry providing detailed descr...
         29    true news text entry providing detailed descri...
         29    false news text entry providing detailed descr...
         12    false news text entry providing detailed descr...
         7     true news text entry providing detailed descri...
         5     false news text entry providing detailed descr...
         27    true news text entry providing detailed descri...
         13    true news text entry providing detailed descri...
         10    true news text entry providing detailed descri...
         17    true news text entry providing detailed descri...
         1     false news text entry providing detailed descr...
         28    true news text entry providing detailed descri...
         6     false news text entry providing detailed descr...
         25    false news text entry providing detailed descr...
         24    true news text entry providing detailed descri...
         25    true news text entry providing detailed descri...
         3     false news text entry providing detailed descr...
         19    true news text entry providing detailed descri...
         18    false news text entry providing detailed descr...
         2     false news text entry providing detailed descr...
         8     false news text entry providing detailed descr...
         0     false news text entry providing detailed descr...
         16    true news text entry providing detailed descri...
         3     true news text entry providing detailed descri...
         17    false news text entry providing detailed descr...
         1     true news text entry providing detailed descri...
         11    true news text entry providing detailed descri...
         20    false news text entry providing detailed descr...
         20    true news text entry providing detailed descri...
         21    false news text entry providing detailed descr...
         13    false news text entry providing detailed descr...
         7     false news text entry providing detailed descr...
         15    false news text entry providing detailed descr...
         12    true news text entry providing detailed descri...
         16    false news text entry providing detailed descr...
         27    false news text entry providing detailed descr...
         14    true news text entry providing detailed descri...
         8     true news text entry providing detailed descri...
         19    false news text entry providing detailed descr...
         26    false news text entry providing detailed descr...
         21    true news text entry providing detailed descri...
         18    true news text entry providing detailed descri...
         4     false news text entry providing detailed descr...
         22    true news text entry providing detailed descri...
         28    false news text entry providing detailed descr...
         5     true news text entry providing detailed descri...
         23    true news text entry providing detailed descri...
         23    false news text entry providing detailed descr...
         11    false news text entry providing detailed descr...
         6     true news text entry providing detailed descri...
         Name: Text, dtype: object
```

```
In [36]: from sklearn.feature_extraction.text import TfidfVectorizer
```

In [37]:
```python
vectorizer = TfidfVectorizer(max_features = 500, lowercase =False, ngram_range = (1,2)
```

In [38]:
```python
x = dataset.iloc[:50,0]
y = dataset.iloc[:50,1]
```

In [39]:
```python
x
```

Out[39]:
```
15    true news text entry providing detailed descri...
10    false news text entry providing detailed descr...
2     true news text entry providing detailed descri...
24    false news text entry providing detailed descr...
4     true news text entry providing detailed descri...
9     true news text entry providing detailed descri...
9     false news text entry providing detailed descr...
22    false news text entry providing detailed descr...
26    true news text entry providing detailed descri...
0     true news text entry providing detailed descri...
14    false news text entry providing detailed descr...
29    true news text entry providing detailed descri...
29    false news text entry providing detailed descr...
12    false news text entry providing detailed descr...
7     true news text entry providing detailed descri...
5     false news text entry providing detailed descr...
27    true news text entry providing detailed descri...
13    true news text entry providing detailed descri...
10    true news text entry providing detailed descri...
17    true news text entry providing detailed descri...
1     false news text entry providing detailed descr...
28    true news text entry providing detailed descri...
6     false news text entry providing detailed descr...
25    false news text entry providing detailed descr...
24    true news text entry providing detailed descri...
25    true news text entry providing detailed descri...
3     false news text entry providing detailed descr...
19    true news text entry providing detailed descri...
18    false news text entry providing detailed descr...
2     false news text entry providing detailed descr...
8     false news text entry providing detailed descr...
0     false news text entry providing detailed descr...
16    true news text entry providing detailed descri...
3     true news text entry providing detailed descri...
17    false news text entry providing detailed descr...
1     true news text entry providing detailed descri...
11    true news text entry providing detailed descri...
20    false news text entry providing detailed descr...
20    true news text entry providing detailed descri...
21    false news text entry providing detailed descr...
13    false news text entry providing detailed descr...
7     false news text entry providing detailed descr...
15    false news text entry providing detailed descr...
12    true news text entry providing detailed descri...
16    false news text entry providing detailed descr...
27    false news text entry providing detailed descr...
14    true news text entry providing detailed descri...
8     true news text entry providing detailed descri...
19    false news text entry providing detailed descr...
26    false news text entry providing detailed descr...
Name: Text, dtype: object
```

In [40]: `y`

Out[40]:
```
15    0
10    1
2     0
24    1
4     0
9     0
9     1
22    1
26    0
0     0
14    1
29    0
29    1
12    1
7     0
5     1
27    0
13    0
10    0
17    0
1     1
28    0
6     1
25    1
24    0
25    0
3     1
19    0
18    1
2     1
8     1
0     1
16    0
3     0
17    1
1     0
11    0
20    1
20    0
21    1
13    1
7     1
15    1
12    0
16    1
27    1
14    0
8     0
19    1
26    1
Name: label, dtype: int64
```

In [41]: **from** sklearn.model_selection **import** train_test_split

In [42]: train_data, test_data, train_label, test_label **=** train_test_split(x,y, test_size **=** 0.5

In [43]: vec_train_data **=** vectorizer.fit_transform(train_data)

In [44]: type(vec_train_data)

Out[44]: scipy.sparse.csr.csr_matrix

In [45]:
```python
vec_train_data = vec_train_data.toarray()
```

In [46]:
```python
type(vec_train_data)
```

Out[46]: numpy.ndarray

In [47]:
```python
vec_test_data = vectorizer.fit_transform(test_data)
```

In [48]:
```python
vec_test_data = vec_test_data.toarray()
```

In [49]:
```python
vec_train_data.shape, vec_test_data.shape
```

Out[49]: ((25, 15), (25, 15))

In [50]: vec_train_data

```
Out[50]: array([[0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
         0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
         0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
        [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
         0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
         0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598],
```

```
       [0.24445936, 0.24445936, 0.24445936, 0.24445936, 0.24445936,
        0.41390568, 0.41390568, 0.24445936, 0.24445936, 0.24445936,
        0.24445936, 0.24445936, 0.24445936, 0.        , 0.        ],
       [0.24812616, 0.24812616, 0.24812616, 0.24812616, 0.24812616,
        0.        , 0.        , 0.24812616, 0.24812616, 0.24812616,
        0.24812616, 0.24812616, 0.24812616, 0.40172598, 0.40172598]])
```

In [51]: `train_data = pd.DataFrame(vec_train_data, columns = vectorizer.get_feature_names())`

In [52]: `test_data = pd.DataFrame(vec_test_data, columns = vectorizer.get_feature_names())`

In [53]: `train_data`

Out[53]:

	description	detailed	detailed description	entry	entry providing	false	false news	news	news text	providing
0	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
1	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
2	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
3	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
4	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
5	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
6	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
7	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
8	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
9	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
10	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
11	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
12	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
13	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
14	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
15	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
16	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
17	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
18	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
19	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
20	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
21	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
22	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126
23	0.244459	0.244459	0.244459	0.244459	0.244459	0.413906	0.413906	0.244459	0.244459	0.244459
24	0.248126	0.248126	0.248126	0.248126	0.248126	0.000000	0.000000	0.248126	0.248126	0.248126

In [54]: `test_data`

Out[54]:

	description	detailed	detailed description	entry	entry providing	false	false news	news	news text	providing
0	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
1	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
2	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
3	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
4	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
5	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
6	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
7	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
8	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
9	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
10	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
11	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
12	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
13	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
14	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
15	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
16	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
17	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
18	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
19	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
20	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
21	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
22	0.240504	0.240504	0.240504	0.240504	0.240504	0.000000	0.000000	0.240504	0.240504	0.240504
23	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535
24	0.251535	0.251535	0.251535	0.251535	0.251535	0.389891	0.389891	0.251535	0.251535	0.251535

In [55]:
```python
from sklearn.naive_bayes import MultinomialNB
```

In [56]:
```python
clf = MultinomialNB()
```

In [57]:
```python
clf.fit(train_data, train_label)
```

Out[57]: MultinomialNB()

In [58]:
```python
y_pred = clf.predict(test_data)
```

```
In [59]:  test_label
```

```
Out[59]:  12    1
          21    1
          8     1
          27    1
          13    0
          19    1
          3     1
          25    0
          16    0
          17    0
          29    1
          4     0
          20    1
          26    0
          24    1
          9     1
          7     1
          14    0
          8     0
          5     1
          0     0
          27    0
          24    0
          17    1
          0     1
          Name: label, dtype: int64
```

```
In [60]:  y_pred
```

```
Out[60]:  array([1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0,
                 0, 1, 1], dtype=int64)
```

```
In [61]:  from sklearn.metrics import accuracy_score
```

```
In [62]:  accuracy_score(test_label, y_pred)
```

```
Out[62]:  1.0
```

```
In [63]:  y_pred_train = clf.predict(train_data)
```

```
In [64]:  accuracy_score(train_label, y_pred_train)
```

```
Out[64]:  1.0
```

```
In [65]:  txt = input("enter the news")
```

          enter the newsThis is the true news text for entry 21, providing detailed descriptio
          n.

```
In [66]:  news = clean_row(str(txt))
```

```
In [69]:  transformed_news = vectorizer.transform([news]).toarray()
          pred = clf.predict(transformed_news)
```

In [70]:
```python
if pred == 0:
    print("News is correct")
else:
    print("News is worng")
```

News is correct

In [71]:
```python
txt = 'This is the true news text for entry 21, providing detailed description.'
```

In [73]:
```python
news = clean_row(str(txt))
```

In [74]:
```python
news
```

Out[74]: 'true news text entry providing detailed description'

In [76]:
```python
pred = clf.predict(vectorizer.transform([news]).toarray())
```

In [77]:
```python
pred
```

Out[77]: array([0], dtype=int64)