

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB, BernoulliNB
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, roc_auc_score, r
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
import plotly.express as px
import string
import warnings
warnings.filterwarnings('ignore')
```

```
In [2]: df=pd.read_csv('fake_news.csv')
```

```
In [3]: df.head()
```

Out[3]:

|   | Unnamed: 0 |   | title   | text   | label |
|---|------------|---|---|--|-------|
| 0 | 0          | LAW ENFORCEMENT ON HIGH ALERT Following Threat... | No comment is expected from Barack Obama Membe... |  | 1     |
| 1 | 1          |   | NaN   | Did they post their votes for Hillary already? | 1     |
| 2 | 2          | UNBELIEVABLE! OBAMA'S ATTORNEY GENERAL SAYS MO... | Now, most of the demonstrators gathered last ...  |  | 1     |
| 3 | 3          | Bobby Jindal, raised Hindu, uses story of Chri... | A dozen politically active pastors came here f... |  | 0     |
| 4 | 4          | SATAN 2: Russia unvelis an image of its terrif... | The RS-28 Sarmat missile, dubbed Satan 2, will... |  | 1     |

```
In [4]: df.drop('Unnamed: 0',axis=1,inplace=True)
```

```
In [5]: df.head()
```

Out[5]:

|   |   | title   | text   | label |
|---|---|---|--|-------|
| 0 | LAW ENFORCEMENT ON HIGH ALERT Following Threat... | No comment is expected from Barack Obama Membe... |  | 1     |
| 1 |   | NaN   | Did they post their votes for Hillary already? | 1     |
| 2 | UNBELIEVABLE! OBAMA'S ATTORNEY GENERAL SAYS MO... | Now, most of the demonstrators gathered last ...  |  | 1     |
| 3 | Bobby Jindal, raised Hindu, uses story of Chri... | A dozen politically active pastors came here f... |  | 0     |
| 4 | SATAN 2: Russia unvelis an image of its terrif... | The RS-28 Sarmat missile, dubbed Satan 2, will... |  | 1     |

```
In [6]: df.isnull().sum().sum()
```

```
Out[6]: 597
```

```
In [7]: df.fillna(' ',inplace=True)
```

```
In [8]: df.tail()
```

```
Out[8]:
```

|       | title   | text  | label |
|-------|---|---|-------|
| 72129 | Russians steal research on Trump in hack of U.... | WASHINGTON (Reuters) - Hackers believed to be ... | 0     |
| 72130 | WATCH: Giuliani Demands That Democrats Apolog...  | You know, because in fantasyland Republicans n... | 1     |
| 72131 | Migrants Refuse To Leave Train At Refugee Camp... | Migrants Refuse To Leave Train At Refugee Camp... | 0     |
| 72132 | Trump tussle gives unpopular Mexican leader mu... | MEXICO CITY (Reuters) - Donald Trump's combati... | 0     |
| 72133 | Goldman Sachs Endorses Hillary Clinton For Pre... | Goldman Sachs Endorses Hillary Clinton For Pre... | 1     |

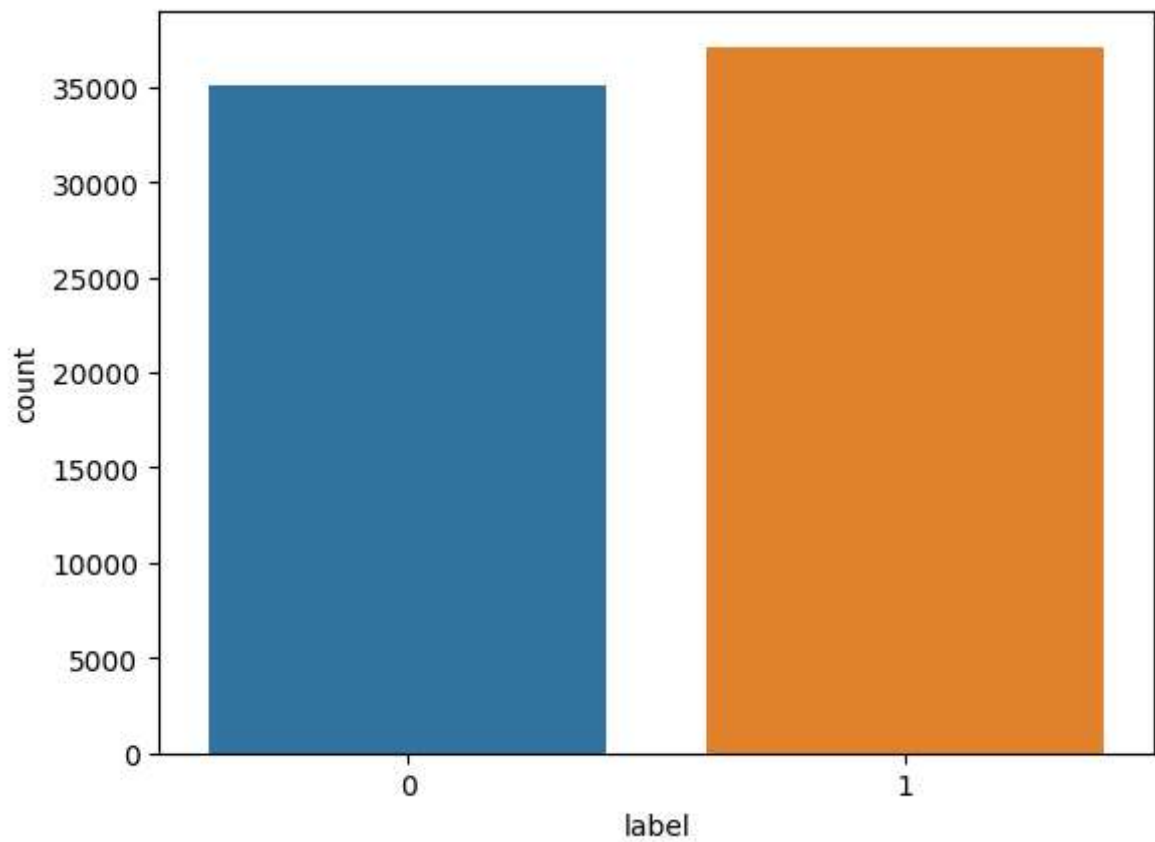
```
In [9]: features = df[['title', 'text']]
labels = df['label']
```

```
In [10]: df.label.value_counts()
```

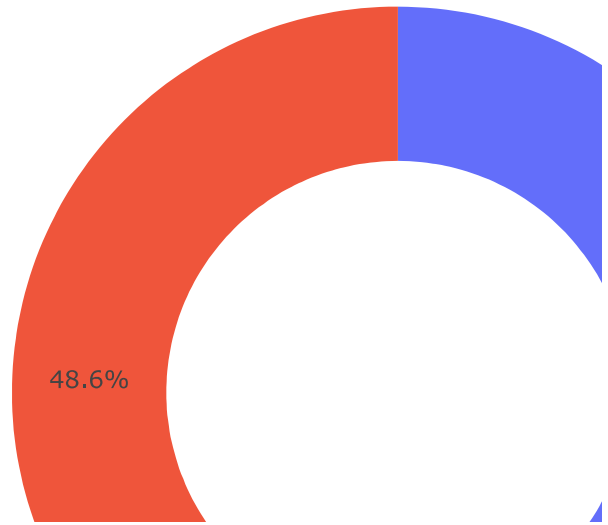
```
Out[10]: 1    37106
0    35028
Name: label, dtype: int64
```

```
In [11]: sns.countplot(x=labels)
```

```
Out[11]: <AxesSubplot: xlabel='label', ylabel='count'>
```



```
In [12]: px.pie(
          df,
          names=df['label'].value_counts().index,
          values=df['label'].value_counts().values,
          hole=.6)
```



```
In [13]: stopwords = set(stopwords.words('english'))
stopword
```

```
Out[13]: {'a',
'about',
'above',
'after',
'again',
'against',
'ain',
'all',
'am',
'an',
'and',
'any',
'are',
'aren',
"aren't",
'as',
'at',
'be',
'because',
...}
```

```
In [ ]:
```

```
In [14]: def tex(text):
remove_pun=[ch for ch in text if ch not in string.punctuation ]
char_joining=''.join(remove_pun)

#stopwords remove
text=[word for word in char_joining.split() if word.lower() not in stopwo
return text
```

```
In [15]: text_column='text'
label_column = 'label'
```

```
In [16]: df[text_column]=df[text_column].apply(tex)
```

```
In [17]: df[text_column]
```

```
Out[17]: 0      [comment, expected, Barack, Obama, Members, FY...
1      [post, votes, Hillary, already]
2      [demonstrators, gathered, last, night, exercis...
3      [dozen, politically, active, pastors, came, pr...
4      [RS28, Sarmat, missile, dubbed, Satan, 2, repl...

...
72129  [WASHINGTON, Reuters, Hackers, believed, worki...
72130  [know, fantasyland, Republicans, never, questi...
72131  [Migrants, Refuse, Leave, Train, Refugee, Camp...
72132  [MEXICO, CITY, Reuters, Donald, Trump's, comba...
72133  [Goldman, Sachs, Endorses, Hillary, Clinton, P...
Name: text, Length: 72134, dtype: object
```

```
In [18]: lemmatizer = WordNetLemmatizer()
```

```
In [19]: def lemmatize_text(text):
    lemmatized_text=' '.join([ lemmatizer.lemmatize(word) for word in text])
    return lemmatized_text

df[text_column] = df[text_column].apply(lemmatize_text)
```

```
In [20]: df.head()
```

Out[20]:

|   | title   | text  | label |
|---|---|---|-------|
| 0 | LAW ENFORCEMENT ON HIGH ALERT Following Threat... | comment expected Barack Obama Members FYF911 F... | 1     |
| 1 |   | post vote Hillary already                         | 1     |
| 2 | UNBELIEVABLE! OBAMA'S ATTORNEY GENERAL SAYS MO... | demonstrator gathered last night exercising co... | 1     |
| 3 | Bobby Jindal, raised Hindu, uses story of Chri... | dozen politically active pastor came private d... | 0     |
| 4 | SATAN 2: Russia unvelis an image of its terrif... | RS28 Sarmat missile dubbed Satan 2 replace SS1... | 1     |

```
In [70]: vectorizer = TfidfVectorizer()
x = vectorizer.fit_transform(df[text_column])
y = df[label_column]
```

```
In [25]: xtrain,xtest,ytrain,ytest=train_test_split(x,y,train_size=.7)
```

```
In [26]: mnb_model = MultinomialNB(alpha=1.0, fit_prior=True, class_prior=None)
mnb_model.fit(xtrain, ytrain)
```

Out[26]:

```
▼ MultinomialNB
MultinomialNB()
```

```
In [46]: df.head(1)
```

Out[46]:

|   | title   | text  | label |
|---|---|---|-------|
| 0 | LAW ENFORCEMENT ON HIGH ALERT Following Threat... | comment expected Barack Obama Members FYF911 F... | 1     |

```
In [27]: from sklearn.metrics import precision_recall_curve, auc
```

```
In [29]: def evaluate_model(model, xtest, ytest):  
    y_pred = model.predict(xtest)  
    accuracy = accuracy_score(ytest, y_pred)  
    cm = confusion_matrix(ytest, y_pred)  
    prob = model.predict_proba(xtest)[: , 1]  
    roc_auc = roc_auc_score(ytest, prob)  
    fpr, tpr, thresholds = roc_curve(ytest, prob)  
    precision, recall, _ = precision_recall_curve(ytest, prob)  
    pr_auc = auc(recall, precision)  
  
    return {  
        'Accuracy': accuracy,  
        'Confusion Matrix': cm,  
    }
```

```
In [30]: nb_results = evaluate_model(mnb_model, xtest, ytest)  
nb_results
```

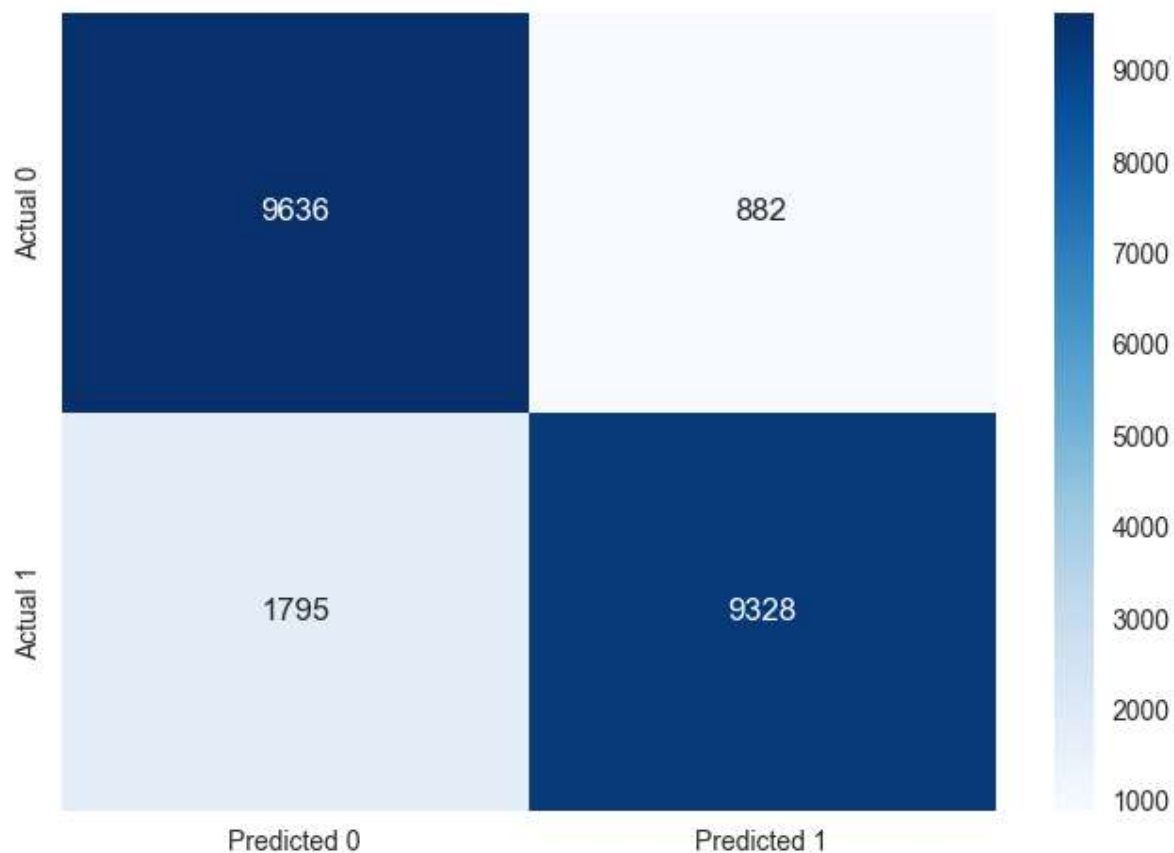
```
Out[30]: {'Accuracy': 0.8762996164687399,  
          'Confusion Matrix': array([[9636,  882],  
                                     [1795, 9328]], dtype=int64)}
```

```
In [32]: cm = confusion_matrix(ytest, mnb_model.predict(xtest))  
cm
```

```
Out[32]: array([[9636,  882],  
               [1795, 9328]], dtype=int64)
```

```
In [33]: sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
                xticklabels=['Predicted 0', 'Predicted 1'],
                yticklabels=['Actual 0', 'Actual 1'])
```

Out[33]: <AxesSubplot: >



```
In [64]: print(classification_report(ytest, mnb_model.predict(xtest)))
```

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.84      | 0.92   | 0.88     | 10518   |
| 1            | 0.91      | 0.84   | 0.87     | 11123   |
| accuracy     |           |        | 0.88     | 21641   |
| macro avg    | 0.88      | 0.88   | 0.88     | 21641   |
| weighted avg | 0.88      | 0.88   | 0.88     | 21641   |



```
In [67]: ran_text = input()

preprocessed_text = tex(ran_text) # remove punctuation
lemmatized_text = lemmatize_text(preprocessed_text) # text scaling
text_vector = vectorizer.transform([lemmatized_text])
```

Migrants Refuse To Leave Train At Refugee Camp In Hungary\n\nThousands of migrants flooded into a train station in the Hungarian capital Thursday after police lifted a two-day blockade, but some who boarded a train they thought was going to Germany ended up instead at a refugee camp just miles from Budapest.\n\nThe Associated Press reports that "excited migrants piled into a newly arrived train at the Keleti station in Hungary\'s capital despite announcements in Hungarian and English that all services from the station to Western Europe had been canceled. A statement on the main departures board said no more trains to Austria or Germany would depart \'due to safety reasons until further notice!\'"\n\n"Many migrants,

In [71]: `preprocessed_text`

```
Out[71]: ['Migrants',
          'Refuse',
          'Leave',
          'Train',
          'Refugee',
          'Camp',
          'HungarynnThousands',
          'migrants',
          'flooded',
          'train',
          'station',
          'Hungarian',
          'capital',
          'Thursday',
          'police',
          'lifted',
          'twoday',
          'blockade',
          'boarded',
          'train',
          'thought',
          'going',
          'Germany',
          'ended',
          'instead',
          'refugee',
          'camp',
          'miles',
          'BudapestnnThe',
          'Associated',
          'Press',
          'reports',
          'excited',
          'migrants',
          'piled',
          'newly',
          'arrived',
          'train',
          'Keleti',
          'station',
          'Hungarys',
          'capital',
          'despite',
          'announcements',
          'Hungarian',
          'English',
          'services',
          'station',
          'Western',
          'Europe',
          'canceled',
          'statement',
          'main',
          'departures',
          'board',
          'said',
          'trains',
```

```
'Austria',
'Germany',
'would',
'depart',
'due',
'safety',
'reasons',
'noticennMany',
'migrants']
```

In [72]: lemmatized\_text

Out[72]: 'Migrants Refuse Leave Train Refugee Camp HungarynnThousands migrant flooded train station Hungarian capital Thursday police lifted twoday blockade boarde d train thought going Germany ended instead refugee camp mile BudapestnnThe A ssociated Press report excited migrant piled newly arrived train Keleti stati on Hungarys capital despite announcement Hungarian English service station We stern Europe canceled statement main departure board said train Austria Germa ny would depart due safety reason noticennMany migrant'

In [73]: text\_vector

Out[73]: <1x363106 sparse matrix of type '<class 'numpy.float64'>' with 50 stored elements in Compressed Sparse Row format>

In [74]: mnb\_model.predict(text\_vector)

Out[74]: array([0], dtype=int64)

In [75]: save\_model(mnb\_model, 'fake\_news\_nlp')

Model Successfully Saved

Out[75]: (MultinomialNB(alpha=1.0, class\_prior=None, fit\_prior=True, force\_alpha='war n'),  
'fake\_news\_nlp.pkl')

In [ ]: