

# **Fake News Detection using NLP techniques**

Fake news detection is a hot topic in the field of natural language processing. In this article, we are using this dataset for news classification using NLP techniques. We are given two input files. One with real news and the other one with fake news. Let us dig further into the given data.

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

real.head

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

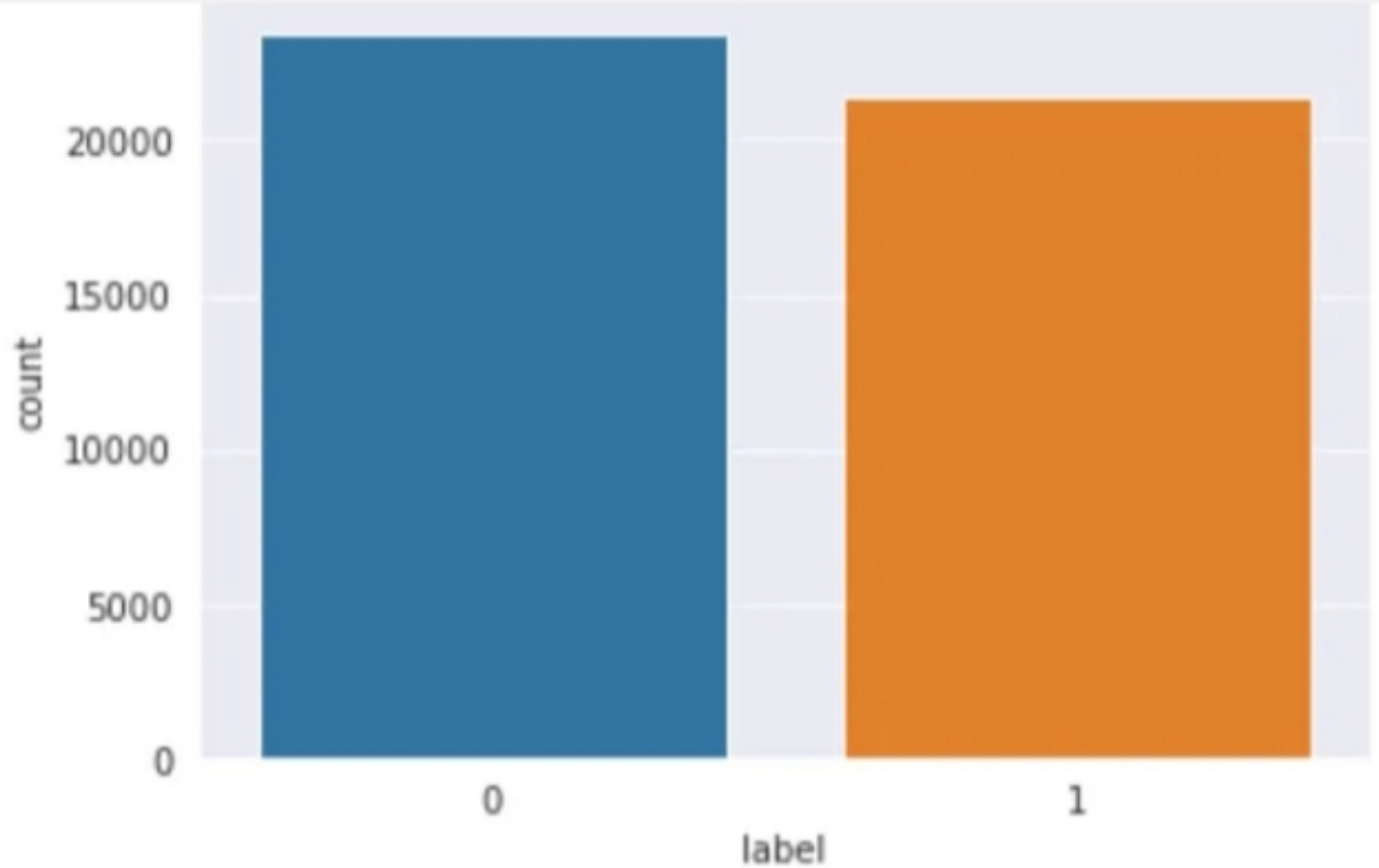
fake.head

Our job is to create a model which predicts whether a given news is real or fake. As this is a supervised learning problem, we are creating a target column named 'label' in both real and fake news data and concatenating them.

```
1 real['label'] = 1
2 fake['label'] = 0
3 data = pd.concat([real, fake])
```

Now we have the input where real news has the value of label as 1 and fake news have the value of label as 0. We have to check whether our data is balanced. We use seaborn library to plot the counts of real and fake news.

```
1  import seaborn as sns
2  sns.set_style("darkgrid")
3  sns.countplot(data['label']);
```



countplot for 'label' column

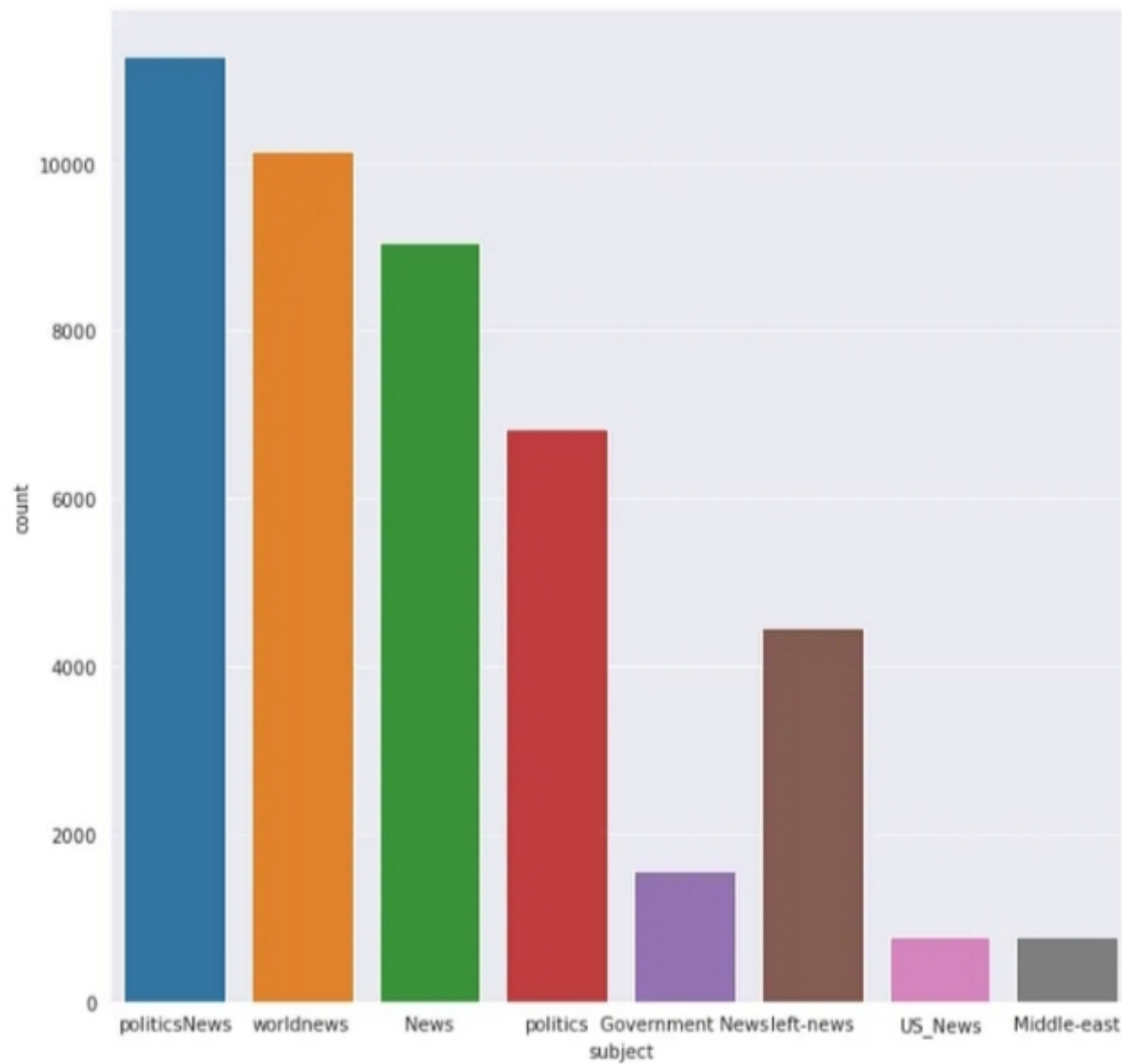
We can conclude from the plot that the data is balanced. This is an important step because there are lot of real world datasets that are imbalanced. Let us check for null values in the data.



```
title      0
text       0
subject    0
date       0
label      0
dtype: int64
```

Luckily, we have no null values. Then, we look at all the columns in the data. There are 5 columns in the data- title, text, subject, date and label. Let us examine the subjects.

```
1 import matplotlib.pyplot as plt
2 data['subject'].value_counts()
3 plt.figure(figsize = (10,10))
4 sns.set_style("darkgrid")
5 sns.countplot(data['subject']);
```

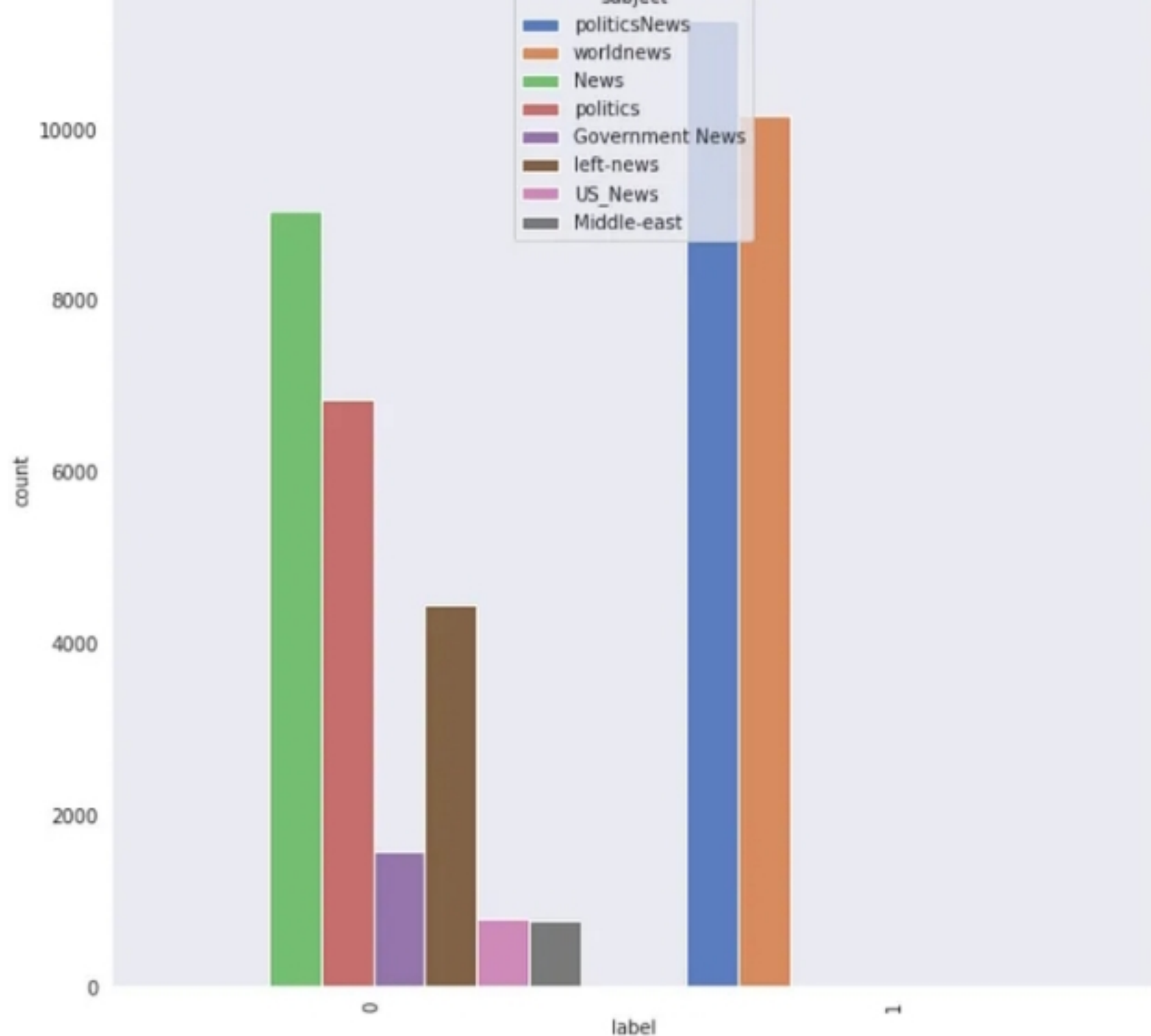


## Countplot of subjects

There are news about 8 subjects.  
We have the largest number of news from politicsNews. Let us dig further into this.

```
1 plt.figure(figsize = (10,10))  
2 sns.set_style("dark")  
3 chart = sns.countplot(x = "label")  
4 chart.set_xticklabels(chart.get_x
```





Contplot of subject with label

It is clear from the plot that all our real news belongs to 2 subjects. That seems to be strange. It might be because our data is taken only from a small period of time. Let us concatenate title and text fields