Data Science news Estimator:

- Created a tool that estimates news which fake or real

Packages: pandas, numpy, sklearn, matplotlib, seaborn.

The solving mechanism

- build machine learning model using python

Describe the dataset

- Data source:
  - https://drive.google.com/file/d/1er9NJTLUA3qnRuyhfzuN0XUsoIC4a-_q/view
- Data description
- I use pandas library to description dataset
  - df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6335 entries, 0 to 6334
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   Unnamed: 0  6335 non-null   int64 
 1   title       6335 non-null   object
 2   text        6335 non-null   object
 3   label       6335 non-null   object
dtypes: int64(1), object(3)
memory usage: 198.1+ KB
```

  - number of rows (6335) and num of columns (4)
  - name of columns and data type for each column
  - their no null values in columns
  - ensure that the dataset is clean

descriptive statistics and data distribution charts

I looked at the distributions of the data and the value counts for the various categorical variables. Below are a few highlights from the pivot tables.

Label	Count
FAKE	2410
REAL	3171

TABLE 1

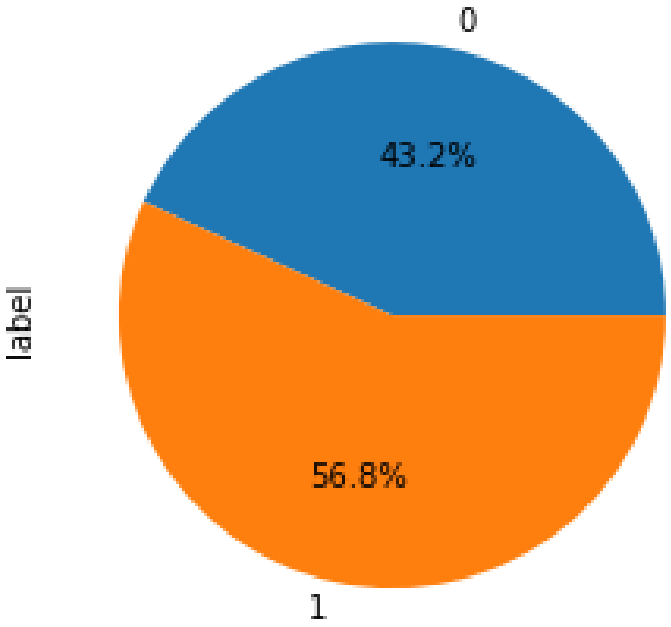

Figure 1

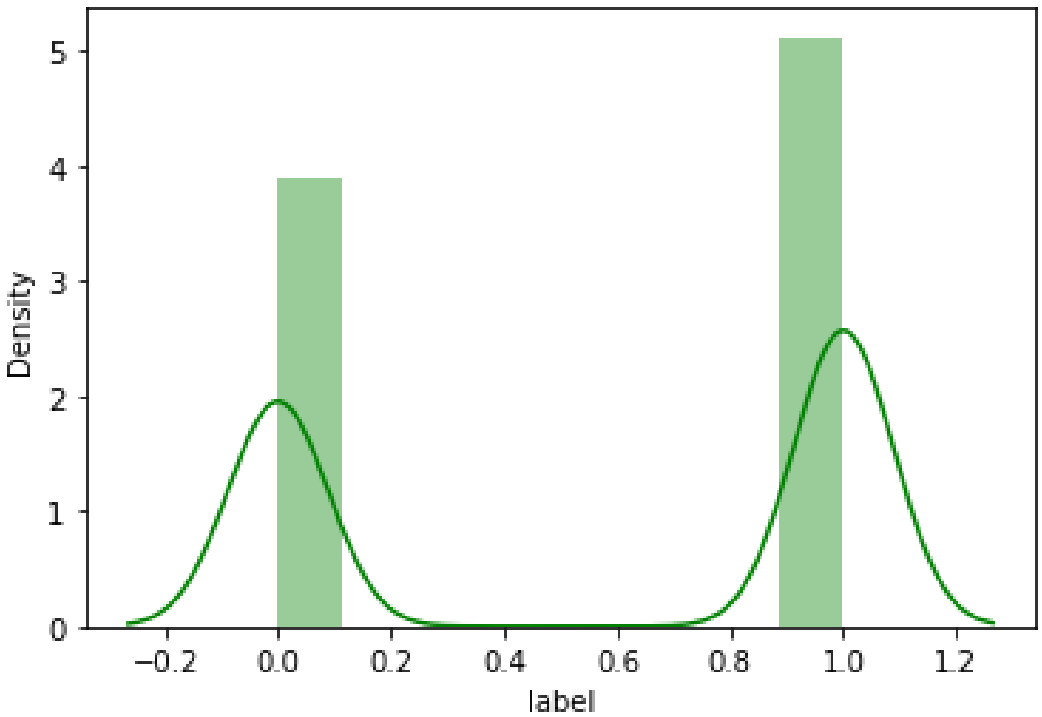

Figure 2

Comment

- (Table 1) pivot_table explain counts of fake and real news
- (figure 1) pie explain percentage of REAL (1) and FAKE (0) news.
- (figure 2) distplot explain distribution of data in dataset

Model Building

First, I split data to X and Y

I also split the X ,Y into train and tests sets with a test size of 20%.

And I apply **TfidfVectorizer** to make transform the X to help us to use it in training and predict

I tried model:

Passive Aggressive Classifier– Belongs to the category of online learning algorithms in machine learning. It works by responding as passive for correct classifications and responding as aggressive for any miscalculation

- # Model performance

The Passive Aggressive Classifier model get performance high.
Passive Aggressive Classifier: = 93.05%