```
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.metrics import accuracy_score
from sklearn.metrics import classification_report
import re
import string


def pre_data():
  real_data=pd.read_csv("/content/True.csv")
  fake_data=pd.read_csv("/content/Fake.csv")
  real_data["class"]=0
  fake_data["class"]=1
  df_merge = pd.concat([real_data, fake_data], axis =0 )
  df = df_merge.drop(["title", "subject","date"], axis = 1)
  df = df.sample(frac = 1)

  df.reset_index(inplace = True)
  df.drop(["index"], axis = 1, inplace = True)

  def wordopt(text):
    text = text.lower()
    text = re.sub('\[.*?\]', '', text)
    text = re.sub("\\W"," ",text)
    text = re.sub('https?://\S+|www\.\S+', '', text)
    text = re.sub('<.*?>+', '', text)
    text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
    text = re.sub('\n', '', text)
    text = re.sub('\w*\d\w*', '', text)
    return text

  df["text"] = df["text"].apply(wordopt)
  return df


data=pre_data()
data
```

	text	class
0	donald trump went on another uncontrollable ra...	1
1	amman reuters sabih al masri jordan s mos...	0
2	washington reuters u s senate judiciary c...	0
3	moscow reuters russian security council se...	0
4	just released hysterical video of sarah palin...	1
...	...	...
44893	we don t know whether the charges against fiat...	1
44894	washington reuters it would be a game cha...	0
44895	thirteen reasons to declare muslim brotherhood...	1
44896	cj pearson the year old conservative social ...	1
44897	in dr kermit gosnell was convicted of killing...	1

44898 rows × 2 columns

```
x=data["text"]
y=data["class"]

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.30)
```

```
def training_data(x_train,y_train):
  from sklearn.feature_extraction.text import TfidfVectorizer

  vectorization = TfidfVectorizer()
  xv_train = vectorization.fit_transform(x_train)

dec_tr,train_df=training_data(x_train,y_train)



def testing_data(dec_tr,train_df):
  pred_dt = dec_tr.predict(train_df)
  return pred_dt
```

```
c=testing_data(dec_tr,train_df)
accuracy=accuracy_score(y_test,c)
accuracy
```

```
0.9949517446176689
```

```
def output_lable(n):
    if n == 0:
        return "Fake News"
    elif n == 1:
        return "Not A Fake News"
```