

Fake news detector:

Importing Packages:

In [1]:

```
import numpy as np
import pandas as pd
import itertools
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.metrics import accuracy_score, confusion_matrix
```

Importing data

In [20]:

```
file = pd.read_csv(r"C:\Users\USP\AppData\Local\Temp\Templ_news.zip\news.csv")

print("This is our data\n")
print(file.shape)
file.head()
```

This is our data

(6335, 4)

Out[20]:

| | Unnamed: 0 | title | text | label |
|---|------------|---|--|-------|
| 0 | 8476 | You Can Smell Hillary's Fear | Daniel Greenfield, a Shillman Journalism Fello... | FAKE |
| 1 | 10294 | Watch The Exact Moment Paul Ryan Committed Pol... | Google Pinterest Digg Linkedin Reddit Stumbleu... | FAKE |
| 2 | 3608 | Kerry to go to Paris in gesture of sympathy | U.S. Secretary of State John F. Kerry said Mon... | REAL |
| 3 | 10142 | Bernie supporters on Twitter erupt in anger ag... | — Kaydee King (@KaydeeKing) November 9, 2016 T... | FAKE |
| 4 | 875 | The Battle of New York: Why This Primary Matters | It's primary day in New York and front-runners... | REAL |

Labelling

In [21]:

```
labelling = file.label
labelling.head()
```

Out[21]:

```
0    FAKE
1    FAKE
2    REAL
3    FAKE
4    REAL
Name: label, dtype: object
```

Splitting data

In [5]:

```
x_train, x_test, y_train, y_test = train_test_split(file['text'], labels, test_size=0.2,
random_state=7)
```

In [6]:

```
x_train
```

Out[6]:

```
6237 The head of a leading survivalist group has ma...
3722 < > Arnaldo Rodgers is a trained and educated ...
5774 Patty Sanchez, 51, used to eat 13,000 calories...
336 But Benjamin Netanyahu's reelection was regard...
3622 John Kasich was killing it with these Iowa vot...
4024 Email \nIt has become increasingly apparent th...
1537 Tesla, 'World's Safest Car,' Explodes Like a B...
5200 For those of you salivating – or trembling – a...
279 American Express disowns Pink Floyd singer Rog...
4973 Osteoporosis (yes, magnesium is more important...
1283 Duterte Calls US Admin 'Monkeys' for Halting A...
4104 (CNN) Donald Trump dominated Super Tuesday, no...
876 Donald Trump has just five days until the Repu...
4769 Home / #Solutions / Colin Kaepernick Just Star...
5080 Black Turnout Soft in Early Voting, Boding Ill...
3601 Seven of the eight leading Republican presiden...
4722 \nOn Sunday evening, side-by-side images of Hi...
3561 Next Prev Swipe left/right This dad's tweets a...
138 Given that Congress has become so utterly dysf...
1949 Fragment of Old Tax Bill Meant to Undercut Mus...
4675 Print \nSen. Bernie Sanders laid out the ways ...
504 By Djuan Wash / filmsforaction.org \nOne canno...
4526 Jeb Bush and Mitt Romney have much in common. ...
4848 We Are Change \n\nWhile many Americans headed ...
4741 NAIROBI, Kenya – President Obama spoke out Sun...
3712 \nFox News reported : \nFive police officers a...
2554 The US' secret drone war has largely been shro...
913 Donald Trump is looking for a veep with the po...
2759 Farms that once generated wealth for entire co...
3131 Email \nIn what Democrats and the Left hoped w...
...
3896 Since we were children, we have been bombarded...
1349 House Republicans investigating the 2012 attac...
3775 During the campaign, Trump had threatened to i...
3116 BNI Store Nov 2 2016 "We are losing control of...
3516 That in itself was a sign that the punch lande...
2298 Most of the headlines from Sen. Elizabeth Warr...
2871 Democratic presidential candidate Hillary Clin...
5195 Speaker At Trump Rally Says He Hopes Hillary C...
391 Archives Michael On Television If Hillary Clin...
6015 With the general election campaign just hours ...
1456 UPDATE: June 4 -- The health insurance enrollm...
1092 The National Security Agency's bulk phone reco...
6118 Abdullah Muhammad al-Muhaysini, a senior al Qa...
3239 Home / News / Fox News Just Exposed Hillary's ...
5000 Levers of Power: Flushing the Vote Down the Me...
2778 (CNN) Donald Trump is now the only Republican ...
4462 License DMCA The American journalist, Edward B...
4441 BREAKING – Investigative Journalist Found Dead...
3656 Genius Kid Trolled White House Halloween Party...
2583 The race for the Democratic presidential nomin...
3257 Now we know why she didn't want those Wall Str...
1372 A big new State Department assessment has iden...
5735 President Barack Obama insisted Thursday that ...
919
4307 By Gordon Duff, Senior Editor on November 2, 2...
5699
2550 It's not that Americans won't elect wealthy pr...
537 Anyone writing sentences like 'nevertheless fu...
1220 More Catholics are in Congress than ever befor...
```

4271 It was hosted by CNN, and the presentation was...
Name: text, Length: 5068, dtype: object

In [7]:

```
x_test
```

Out[7]:

```
3534 A day after the candidates squared off in a fi...
6265 VIDEO : FBI SOURCES SAY INDICTMENT LIKELY FOR ...
3123 It's debate season, where social media has bro...
3940 Mitch McConnell has decided to wager the Repub...
2856 Donald Trump, the actual Republican candidate ...
3031 Not all the claims in the vice presidential de...
4854 WASHINGTON—Top administration officials forcef...
5861 Top Dems want White House to call off Part B d...
307 Corrections and clarifications: An earlier ver...
2956 posted by Eddie Censorship by Facebook has bec...
4594 Email \n\nWSJ The political organization of Vi...
5268 On Sunday during an appearance on "Face the Na...
3929 After Russian President Vladimir Putin complim...
4276 Senate supporters of the Keystone XL pipeline,...
5655 The Republican presidential race continues to ...
5355 Wed, 26 Oct 2016 17:26 UTC © Sandro Perozzi/AP...
2268 (CNN) Matt Bevin, the controversial Kentuckian...
4525 Washington (CNN) He cut billions in taxes, int...
5979 Hillary Clinton won at least four of five prim...
1854 Getty - Saul Loeb/AFP The Wildfire is an opini...
2463 Posted on October 28, 2016 by DCG | 2 Comments...
4460 In the days after King's assassination, Americ...
305 Ohio governor John Kasich, who is still runnin...
4736 Financial Markets , Gold , Market Manipulation...
5098 Sen. Rand Paul, R-Ky., announced today that he...
351 The federal government must make regular inter...
4932 That's one way to quickly characterize the dif...
5477 Posted on October 27, 2016 \nThe editorial boa...
5370 \nThe impact of the Syrian refugee crisis has ...
1328 Reinventing Democracy in America Starts by Vot...
...
1535 Email \nI was in London last weekend to view a...
1390 Truth Revolt October 26, 2016 \nAs usual, Pat ...
444 Hegel remarks somewhere that all great world-h...
640 A Texas jury reached a guilty verdict in the m...
4800 \nCBS NEWS \nNEW YORK -- CBS News has learned ...
6196 The report also discloses new details relevant...
1208 Posted on November 7, 2016 by Carl Herman \nJo...
459 (Live Streams Available Below) \nAnti-Trump pr...
1598 Romney's condemnation, made here at the Stein ...
5622 Wikileaks: Bill Clinton BOASTS of Hillary's 'W...
1955 Email \nTo view photojournalist Orin Langelles'...
1153 CHARLESTON, S.C. — Emanuel AME Church swung op...
3202 The USA Era (1945 - 2008) in Retrospect It wor...
6089 November 1, 2016 UN Chief urges South Africa t...
587 How Western Media Teleported a Child 'Victim' ...
625 Next Swipe left/right Make music great again, ...
1675 The Iraqi Air Force has bombed the convoy of I...
3961 In Texas, Alabama and elsewhere a number of cl...
2646 Get short URL 0 44 0 0 Foreign Minister Sergei...
5544 (CNN) First, DNA on a pizza crust led authorit...
1181 Protesters across the US were on Friday gearin...
4568 UK citizens and war heroes get cheap pre-fab h...
1953 On Friday morning, a US airstrike killed Abu A...
735 law , economy , society , standard of living ,...
1137 Dems File Complaint w/DOJ Against FBI for Inve...
4986 Washington (CNN) President Barack Obama announ...
5789 The revival of middle-class jobs has been one ...
4338 "I can guarantee that," Obama answered when as...
5924 Videos 30 Civilians Die In US Airstrike Called...
6030 The retired neurosurgeon lashed out Friday mor...
Name: text, Length: 1267, dtype: object
```

In [8]:

```
y_train
```

Out[8]:

```
6237    FAKE
3722    FAKE
5774    FAKE
336     REAL
3622    REAL
4024    FAKE
1537    FAKE
5200    REAL
279     FAKE
4973    FAKE
1283    FAKE
4104    REAL
876     REAL
4769    FAKE
5080    FAKE
3601    REAL
4722    FAKE
3561    FAKE
138     REAL
1949    FAKE
4675    FAKE
504     FAKE
4526    REAL
4848    FAKE
4741    REAL
3712    FAKE
2554    FAKE
913     REAL
2759    REAL
3131    FAKE
...
3896    FAKE
1349    REAL
3775    REAL
3116    FAKE
3516    REAL
2298    REAL
2871    REAL
5195    FAKE
391     FAKE
6015    REAL
1456    REAL
1092    REAL
6118    FAKE
3239    FAKE
5000    FAKE
2778    REAL
4462    FAKE
4441    FAKE
3656    FAKE
2583    REAL
3257    REAL
1372    REAL
5735    REAL
919     FAKE
4307    FAKE
5699    FAKE
2550    REAL
537     REAL
1220    REAL
4271    REAL
```

Name: label, Length: 5068, dtype: object

In [9]:

```
y_test
```

Out [9]:

```
3534    REAL
6265    FAKE
3123    REAL
3940    REAL
2856    REAL
3031    REAL
4854    REAL
5861    REAL
307     REAL
2956    FAKE
4594    FAKE
5268    REAL
3929    REAL
4276    REAL
5655    REAL
5355    FAKE
2268    REAL
4525    REAL
5979    REAL
1854    FAKE
2463    FAKE
4460    REAL
305     REAL
4736    FAKE
5098    REAL
351     REAL
4932    REAL
5477    FAKE
5370    FAKE
1328    FAKE
...
1535    FAKE
1390    FAKE
444     FAKE
640     REAL
4800    FAKE
6196    REAL
1208    FAKE
459     FAKE
1598    REAL
5622    FAKE
1955    FAKE
1153    REAL
3202    FAKE
6089    FAKE
587     FAKE
625     FAKE
1675    REAL
3961    REAL
2646    FAKE
5544    REAL
1181    REAL
4568    FAKE
1953    REAL
735     FAKE
1137    FAKE
4986    REAL
5789    REAL
4338    REAL
5924    FAKE
6030    REAL
Name: label, Length: 1267, dtype: object
```

Working with TfidfVectorizer

In [22]:

```
tfidf_vector=TfidfVectorizer(stop_words='english', max_df=0.7)
```

```
tfidf_train=tfidf_vector.fit_transform(x_train)
tfidf_test=tfidf_vector.transform(x_test)
```

In [23]:

```
print(tfidf_train)
```

```
(0, 25318) 0.06356197880227954
(0, 31796) 0.032835312190384304
(0, 53256) 0.08468714610538439
(0, 24260) 0.02591230211547521
(0, 49636) 0.04842197282925203
(0, 3378) 0.044440694689591796
(0, 57236) 0.04731169310146364
(0, 51968) 0.04086068970720691
(0, 23825) 0.06665750504864855
(0, 8860) 0.06997151470751813
(0, 48721) 0.07298886000547623
(0, 10056) 0.08875931355939351
(0, 56878) 0.22907566353089
(0, 8953) 0.07617756364829195
(0, 17858) 0.08179789731649102
(0, 57811) 0.05018134592598926
(0, 23481) 0.13387265628974546
(0, 18378) 0.16712933698409224
(0, 9851) 0.026344757209787056
(0, 27283) 0.114537831765445
(0, 4259) 0.1252399574231224
(0, 19585) 0.06609677424160539
(0, 58158) 0.03931037043190289
(0, 8952) 0.06669214616557032
(0, 15390) 0.18274094576335073
: :
(5067, 25194) 0.06815122596851944
(5067, 40823) 0.06020842915278987
(5067, 51237) 0.06880620615191042
(5067, 49605) 0.06329472428436525
(5067, 37290) 0.07103104644053512
(5067, 55383) 0.07547446468030639
(5067, 42708) 0.06485466744083908
(5067, 42219) 0.14048090961920684
(5067, 13717) 0.06308824962480555
(5067, 12961) 0.12923678067290478
(5067, 56698) 0.06753166786457682
(5067, 26810) 0.08403681959997857
(5067, 58513) 0.07188025218175627
(5067, 18443) 0.06694389628192306
(5067, 40942) 0.06950090572888828
(5067, 50907) 0.08538649936034741
(5067, 54347) 0.076094022784249
(5067, 19852) 0.08868305788802776
(5067, 12311) 0.0807402610722982
(5067, 15063) 0.09662585470375731
(5067, 17502) 0.09662585470375731
(5067, 57890) 0.09662585470375731
(5067, 7909) 0.10127209299180652
(5067, 23082) 0.10127209299180652
(5067, 36693) 0.10127209299180652
```

In [24]:

```
print(tfidf_test)
```

```
(0, 60731) 0.05899712902382916
(0, 60684) 0.033385466151529625
(0, 60271) 0.04581143542258741
(0, 60261) 0.07937859313949312
(0, 59116) 0.10997273171965094
(0, 59036) 0.08042180974421559
(0, 58654) 0.07128159375531905
```

```
(0, 58335) 0.0678398429566027
(0, 57086) 0.12429244186413906
(0, 55170) 0.20939665348422057
(0, 54706) 0.035492943055135416
(0, 54394) 0.10596727423829927
(0, 54238) 0.06234899619642803
(0, 53749) 0.04942070163765446
(0, 53518) 0.117529167732626
(0, 53144) 0.059982387365669215
(0, 52555) 0.14065295472127948
(0, 52483) 0.07360851972393109
(0, 51960) 0.14065295472127948
(0, 51955) 0.030414611451489323
(0, 51663) 0.12057034351821985
(0, 51527) 0.08134873077710283
(0, 51159) 0.053278054236854326
(0, 51005) 0.07360851972393109
(0, 50068) 0.06481045119580665
: :
(1266, 16835) 0.08080116269909657
(1266, 16385) 0.0655727278454052
(1266, 15999) 0.03279926109441314
(1266, 14890) 0.05544599110567492
(1266, 13110) 0.06697165573835866
(1266, 12885) 0.05404331760933019
(1266, 12792) 0.052054416965658004
(1266, 11999) 0.09393931399068145
(1266, 11425) 0.061215890922102396
(1266, 11275) 0.03975222944194533
(1266, 10398) 0.03928759838998747
(1266, 10334) 0.052485971705883705
(1266, 9807) 0.47248142728953435
(1266, 9082) 0.06889577674672712
(1266, 8639) 0.09860931312297018
(1266, 7209) 0.05947909473390904
(1266, 7127) 0.12219547760308587
(1266, 7021) 0.09258263295223065
(1266, 6653) 0.05532410291136702
(1266, 6288) 0.04255750958351483
(1266, 5568) 0.04067869750578144
(1266, 5070) 0.08544954436342893
(1266, 4297) 0.09929470490549547
(1266, 2510) 0.04365670001733164
(1266, 2365) 0.06671538208893119
```

Initializing the PassiveAgressiveClassifier

In [16]:

```
pac=PassiveAggressiveClassifier(max_iter=50)
pac.fit(tfidf_train,y_train)
```

```
y_pred=pac.predict(tfidf_test)
score=accuracy_score(y_test,y_pred)
print(f'Accuracy: {round(score*100,2)}%')
```

Accuracy: 92.82%

The accuracy is 92.82% in this model. To gain insights about the number of False and True negatives and positives, let's construct out confusionmatrix.

In [17]:

```
confusion_matrix(y_test,y_pred, labels=['FAKE','REAL'])
```

Out[17]:

```
array([[588, 50],
       [ 41, 588]], dtype=int64)
```

