In [168]:
```python
import pandas as pd
from collections import Counter
import re
import numpy as np
from sklearn.utils import shuffle
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.stop_words import ENGLISH_STOP_WORDS
from sklearn.metrics import f1_score, accuracy_score , recall_score , pre
import matplotlib.pyplot as plt
from xgboost import XGBClassifier
from sklearn.ensemble import RandomForestClassifier

from sklearn import datasets, linear_model

from matplotlib import pyplot as plt

%matplotlib inline
%config InlineBackend.figure_format = 'retina'
```

In [169]:
```python
# Reading buzzfeedreal dataset
df_buzzr = pd.read_csv("/Users/sonia/Downloads/finalyearproject/Realnews_
```

In [170]:
```python
df_buzzr.columns
```

Out[170]: Index(['Unnamed: 0', 'top_img', 'text', 'authors__001', 'authors__002'
        ,
           'authors__003', 'meta_data__generator', 'authors__004', 'author
        s__005',
           'authors__006', 'authors__007', 'authors__008', 'authors__009',
           'authors__010', 'meta_data__fb_title', 'meta_data__description'
        ,
           'meta_data__title', 'meta_data__modernizr-version',
           'meta_data__foundation-version', 'meta_data__jquery-version',
           'meta_data__pubdate', 'meta_data__copyright', 'meta_data__autho
        r',
           'meta_data__vr__category', 'meta_data__vr__type',
           'meta_data__og__site_name', 'meta_data__og__description',
           'meta_data__og__pubdate', 'meta_data__og__title',
           'meta_data__og__locale', 'meta_data__og__updated_time',
           'meta_data__og__url', 'meta_data__og__image', 'meta_data__og__f
        b_appid',
           'meta_data__og__type', 'meta_data__section', 'meta_data__referr

```
er',
       'meta_data__twitter__description', 'meta_data__twitter__title',
       'meta_data__twitter__url', 'meta_data__twitter__image',
       'meta_data__twitter__creator', 'meta_data__object-hash',
       'meta_data__host', 'meta_data__twitter__site',
       'meta_data__twitter__app__|', 'meta_data__twitter__app__|__ipad
',
       'meta_data__twitter__app__|__iphone',
       'meta_data__twitter__app__|__googleplay', 'meta_data__twitter__
card',
       'meta_data__robots', 'meta_data__apple-itunes-app',
       'meta_data__DC.date.issued', 'meta_data__publish_date',
       'meta_data__al__|__url', 'meta_data__Date', 'meta_data__fb__adm
ins',
       'meta_data__fb__app_id', 'meta_data__date',
       'meta_data__msapplication-TileImage', 'meta_data__application-n
ame',
       'meta_data__msApplication-PackageFamilyName',
       'meta_data__article__publisher', 'meta_data__theme-color',
       'meta_data__fb__pages', 'meta_data__thumbnail',
       'meta_data__template-top', 'meta_data__vr__canonical',
       'meta_data__build', 'meta_data__article__section',
       'meta_data__article__tag', 'meta_data__article__published_time'
,
       'meta_data__article__modified_time',
       'meta_data__google-site-verification', 'meta_data__verify-v1',
       'meta_data__Last-Modified', 'meta_data__keywords', 'meta_data__
ov__id',
       'meta_data__lastmod', 'meta_data__article__author',
       'meta_data__news_keywords', 'canonical_link', 'title', 'url',
       'publish_date__$date', 'source', 'fakeness'],
     dtype='object')
```

In [171]:
```python
#Reading the politifact real dataset
df_polir = pd.read_csv("/Users/sonia/Downloads/finalyearproject/Realnews_
```

In [172]:
```
df_polir.columns
```

Out[172]:
```
Index(['Unnamed: 0', 'top_img', 'text', 'authors__001', 'meta_data__pu
blisher',
       'meta_data__shareaholic__site_name', 'meta_data__shareaholic__l
anguage',
       'meta_data__shareaholic__url',
       'meta_data__shareaholic__article_author_name',
       'meta_data__shareaholic__image',
       ...
       'meta_data__viewport', 'meta_data__news_keywords', 'meta_data__
title',
       'meta_data__modernizr-version', 'canonical_link', 'title', 'url
',
       'publish_date__$date', 'source', 'fakeness'],
      dtype='object', length=109)
```

In [173]:
```
#reading the buzzfeed fake dataset
fakebf=pd.read_csv("/Users/sonia/Downloads/finalyearproject/FakeNews_Clea
```

In [174]:
```
fakebf.columns
```

Out[174]:
```
Index(['Unnamed: 0', 'top_img', 'text', 'authors__001', 'meta_data__pu
blisher',
       'authors__002', 'authors__003', 'authors__004', 'authors__005',
       'authors__006', 'authors__007', 'meta_data__description',
       'meta_data__generator', 'meta_data__author', 'meta_data__og__si
te_name',
       'meta_data__og__description', 'meta_data__og__title',
       'meta_data__og__locale', 'meta_data__og__image',
       'meta_data__og__updated_time', 'meta_data__og__url',
       'meta_data__og__type', 'meta_data__twitter__description',
       'meta_data__twitter__creator', 'meta_data__twitter__url',
       'meta_data__twitter__image', 'meta_data__twitter__title',
       'meta_data__fb__admins', 'meta_data__twitter__site',
       'meta_data__twitter__card', 'meta_data__robots',
       'meta_data__DC.date.issued', 'meta_data__fb__app_id',
       'meta_data__fb__pages', 'meta_data__article__publisher',
       'meta_data__article__author', 'meta_data__keywords',
       'meta_data__google-site-verification', 'meta_data__article__sec
tion',
       'meta_data__article__tag', 'meta_data__article__published_time'
',
       'meta_data__article__modified_time', 'meta_data__viewport',
       'canonical_link', 'title', 'url', 'publish_date__$date', 'sourc
e',
       'fakeness'],
      dtype='object')
```

```
In [175]:  #reading the Politifact fake dataset
           fakepf=pd.read_csv("/Users/sonia/Downloads/finalyearproject/FakeNews_Clea
```

```
In [176]:  fakepf.columns
```
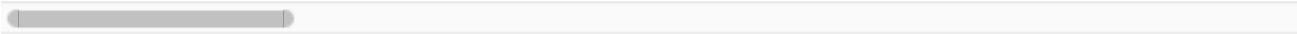
```
Out[176]:  Index(['Unnamed: 0', 'top_img', 'text', 'meta_data__shareaholic__site_
           name',
                   'meta_data__shareaholic__language', 'meta_data__shareaholic__ur
           l',
                   'meta_data__shareaholic__article_author_name',
                   'meta_data__shareaholic__image', 'meta_data__shareaholic__site_
           id',
                   'meta_data__shareaholic__article_published_time',
                   ...
                   'meta_data__author', 'meta_data__viewport', 'canonical_link',
                   'images__-', 'title', 'url', 'movies__-', 'publish_date__$date'
           ,
                   'source', 'fakeness'],
                 dtype='object', length=102)
```

In [177]: `df_buzzr.head(5)`

Out[177]:

	Unnamed: 0	top_img	text	authors__001	authc
0	0	http://a.abcnews.com/images/Politics/AP_donald...	Less than a day after protests over the police...	More Candace	Adar
1	4	http://rightwingnews.com/wp-content/uploads/20...	Obama To UN: 'Giving Up Liberty, Enhances Secu...	Cassy Fiano	
2	26	http://static.politico.com/e9/11/6144cdc24e319...	Getty Images Wealth Of Nations Trump vs. Clint...	Jack Shafer	Eric
3	29	http://a.abcnews.com/images/US/AP_Obama_BM_201...	President Obama today vetoed a bill that would...	John Parkinson	M
4	33	http://rightwingnews.com/wp-content/uploads/20...	CHAOS! NC Protest MOB Ambushes Female Truck Dr...	Cassy Fiano	

5 rows × 87 columns

```python
In [178]: df_buzzr=df_buzzr.drop(columns=['authors__002',
             'authors__003', 'meta_data__generator', 'authors__004', 'authors__
             'authors__006', 'authors__007', 'authors__008', 'authors__009',
             'authors__010', 'meta_data__fb_title', 'meta_data__description',
             'meta_data__title', 'meta_data__modernizr-version',
             'meta_data__foundation-version', 'meta_data__jquery-version',
             'meta_data__pubdate', 'meta_data__copyright', 'meta_data__author',
             'meta_data__vr__category', 'meta_data__vr__type',
             'meta_data__og__site_name', 'meta_data__og__description',
             'meta_data__og__pubdate', 'meta_data__og__title',
             'meta_data__og__locale', 'meta_data__og__updated_time',
             'meta_data__og__url', 'meta_data__og__image', 'meta_data__og__fb_a
             'meta_data__og__type', 'meta_data__section', 'meta_data__referrer'
             'meta_data__twitter__description', 'meta_data__twitter__title',
             'meta_data__twitter__url', 'meta_data__twitter__image',
             'meta_data__twitter__creator', 'meta_data__object-hash',
             'meta_data__host', 'meta_data__twitter__site',
             'meta_data__twitter__app__|', 'meta_data__twitter__app__|__ipad',
             'meta_data__twitter__app__|__iphone',
             'meta_data__twitter__app__|__googleplay', 'meta_data__twitter__car
             'meta_data__robots', 'meta_data__apple-itunes-app',
             'meta_data__DC.date.issued', 'meta_data__publish_date',
             'meta_data__al__|__url', 'meta_data__Date', 'meta_data__fb__admins
             'meta_data__fb__app_id', 'meta_data__date',
             'meta_data__msapplication-TileImage', 'meta_data__application-name
             'meta_data__msApplication-PackageFamilyName',
             'meta_data__article__publisher', 'meta_data__theme-color',
             'meta_data__fb__pages', 'meta_data__thumbnail',
             'meta_data__template-top', 'meta_data__vr__canonical',
             'meta_data__build', 'meta_data__article__section',
             'meta_data__article__tag', 'meta_data__article__published_time',
             'meta_data__article__modified_time',
             'meta_data__google-site-verification', 'meta_data__verify-v1',
             'meta_data__Last-Modified', 'meta_data__keywords', 'meta_data__ov_
             'meta_data__lastmod', 'meta_data__article__author'])
```

In [179]: `df_buzzr`

Out[179]:

	Unnamed: 0	top_img	text	authors__0
0	0	http://a.abcnews.com/images/Politics/AP_donald...	Less than a day after protests over the police...	Mc Canda
1	4	http://rightwingnews.com/wp-content/uploads/20...	Obama To UN: 'Giving Up Liberty, Enhances Secu...	Cassy Fia
2	26	http://static.politico.com/e9/11/6144cdc24e319...	Getty Images Wealth Of Nations Trump vs. Clint...	Jack Sha
			President Obama	

In [180]: `df_buzzr=df_buzzr.drop(columns=['Unnamed: 0', 'top_img'])` *#dropping unnec*

In [181]: `df_buzzr=df_buzzr.drop(columns=['meta_data__news_keywords'])` *#dropping un*

In [182]: `df_buzzr`

9	I've watched every presidential debate ever br...	Jack Shafer	http://www.politico.com/magazine/story/2016/09...	Is Hillary in a No-Win Situation?
10	Story highlights "This, though, is certain: to...	Mj Lee	http://www.cnn.com/2016/09/21/politics/hillary...	Hillary Clinton on police shootings: 'too many...
11	"I take great pride in the fact that I am actu...	Jack Shafer	http://www.politico.com/story/2016/09/gary-joh...	Johnson: 'I'm going to sleep well at the end o...
12	The U.S. Attorney's Office in Manhattan is inv...	More Josh	http://abcnews.go.com/Politics/anthony-weiner-...	Anthony Weiner Under Federal Investigation for...
13	11.7k SHARES SHARE THIS	Grant Stern	http://occupydemocrats.com/2016/09/23/senate-...	Senate Dems Move To Force Trump

```
In [183]: df_polir=df_polir.drop(columns=['authors__002',
          'authors__003', 'meta_data__generator', 'authors__004', 'authors__
          'authors__006', 'authors__007', 'authors__008', 'authors__009',
           'meta_data__fb_title', 'meta_data__description',
          'meta_data__title', 'meta_data__modernizr-version',
          'meta_data__pubdate', 'meta_data__copyright', 'meta_data__author',
          'meta_data__vr__category', 'meta_data__vr__type',
          'meta_data__og__site_name', 'meta_data__og__description',
          'meta_data__og__pubdate', 'meta_data__og__title',
          'meta_data__og__locale', 'meta_data__og__updated_time',
          'meta_data__og__url', 'meta_data__og__image', 'meta_data__og__fb_a
          'meta_data__og__type', 'meta_data__section', 'meta_data__referrer'
          'meta_data__twitter__description', 'meta_data__twitter__title',
          'meta_data__twitter__url', 'meta_data__twitter__image',
          'meta_data__twitter__creator',
          'meta_data__host', 'meta_data__twitter__site',
          'meta_data__twitter__app__|', 'meta_data__twitter__app__|__ipad',
          'meta_data__twitter__app__|__iphone',
          'meta_data__twitter__app__|__googleplay', 'meta_data__twitter__car
          'meta_data__robots',
          'meta_data__DC.date.issued', 'meta_data__publish_date',
          'meta_data__al__|__url', 'meta_data__Date', 'meta_data__fb__admins
          'meta_data__fb__app_id', 'meta_data__date',
          'meta_data__msapplication-TileImage', 'meta_data__application-name
          'meta_data__msApplication-PackageFamilyName',
          'meta_data__article__publisher', 'meta_data__theme-color',
          'meta_data__fb__pages', 'meta_data__thumbnail',
          'meta_data__template-top', 'meta_data__vr__canonical',
          'meta_data__build', 'meta_data__article__section',
          'meta_data__article__tag', 'meta_data__article__published_time',
          'meta_data__article__modified_time',
          'meta_data__google-site-verification', 'meta_data__verify-v1',
          'meta_data__Last-Modified', 'meta_data__keywords', 'meta_data__ov_
          'meta_data__lastmod', 'meta_data__article__author'])
```

In [184]: df_polir

Out[184]:

	Unnamed: 0	top_img	text	authors__001
0	0	http://rightwingnews.com/wp-content/uploads/20...	Famous dog killed in spot she waited a year fo...	missing
1	22	http://i2.cdn.cnn.com/cnnnext/dam/assets/16091...	Story highlights The House Oversight panel vot...	Tom Lobianco
2	43	http://newsbake.com/wp-content/uploads/2016/05...	We are absolutely heartbroken to hear about th...	Nancy Wells
			Nine years ago, a	

In [185]:
```python
df_polir.columns
```

Out[185]:
```
Index(['Unnamed: 0', 'top_img', 'text', 'authors__001', 'meta_data__pu
blisher',
       'meta_data__shareaholic__site_name', 'meta_data__shareaholic__l
anguage',
       'meta_data__shareaholic__url',
       'meta_data__shareaholic__article_author_name',
       'meta_data__shareaholic__image', 'meta_data__shareaholic__site_
id',
       'meta_data__shareaholic__article_published_time',
       'meta_data__shareaholic__shareable_page',
       'meta_data__shareaholic__article_modified_time',
       'meta_data__shareaholic__keywords',
       'meta_data__shareaholic__wp_version', 'meta_data__msApplication
-ID',
       'meta_data__googlebot', 'meta_data__og__image__width',
       'meta_data__og__image__identifier', 'meta_data__og__image__type
',
       'meta_data__og__image__height', 'meta_data__wp-parsely_version'
,
       'meta_data__theme-version', 'meta_data__al__|',
       'meta_data__al__|__app_store_id', 'meta_data__al__|__app_name',
       'meta_data__msvalidate.01', 'meta_data__al__|__package',
       'meta_data__propeller', 'meta_data__viewport',
       'meta_data__news_keywords', 'canonical_link', 'title', 'url',
       'publish_date__$date', 'source', 'fakeness'],
      dtype='object')
```

In [186]:
```python
df_polir=df_polir.drop(columns=['Unnamed: 0', 'top_img','meta_data__publi
               'meta_data__shareaholic__site_name', 'meta_data__shareaholic__lang
               'meta_data__shareaholic__url',
               'meta_data__shareaholic__article_author_name',
               'meta_data__shareaholic__image', 'meta_data__shareaholic__site_id'
               'meta_data__shareaholic__article_published_time',
               'meta_data__shareaholic__shareable_page',
               'meta_data__shareaholic__article_modified_time',
               'meta_data__shareaholic__keywords',
               'meta_data__shareaholic__wp_version', 'meta_data__msApplication-ID
               'meta_data__googlebot', 'meta_data__og__image__width',
               'meta_data__og__image__identifier', 'meta_data__og__image__type',
               'meta_data__og__image__height', 'meta_data__wp-parsely_version',
               'meta_data__theme-version', 'meta_data__al__|',
               'meta_data__al__|__app_store_id', 'meta_data__al__|__app_name',
               'meta_data__msvalidate.01', 'meta_data__al__|__package',
               'meta_data__propeller', 'meta_data__viewport'])
```

In [187]:
```python
df_polir=df_polir.drop(columns=['meta_data__news_keywords'])#dropping unn
```

In [188]: `df_polir`

25	Reps. John Lewis (D-Ga.), Brenda Lawrence (D-M...	John Bresnahan	http://www.politico.com/story/2016/09/congress...	Bla Cauc marches DOJ prote shootin
26	55.6k SHARES SHARE THIS STORY\n\nOnce again, D...	Colin Taylor	http://occupydemocrats.com/2016/09/20/trump-si...	Trump Sile As Poli Credit A Si Immigran
27	"He has a long record of engaging in racist be...	Jack Shafer	http://www.politico.com/story/2016/09/clinton-...	Clinton Trump birther clai A 'racist l
28	In total, 37 states and the District of Columb...	Jack Shafer	http://www.politico.com/story/2016/09/early-vo...	When do early voti start in eve stat
				Rep. Pet

In [189]:
```python
fakebf=fakebf.drop(columns=['meta_data__publisher',
    'authors__002', 'authors__003', 'authors__004', 'authors__005',
    'authors__006', 'authors__007', 'meta_data__description',
    'meta_data__generator', 'meta_data__author', 'meta_data__og__site_
    'meta_data__og__description', 'meta_data__og__title',
    'meta_data__og__locale', 'meta_data__og__image',
    'meta_data__og__updated_time', 'meta_data__og__url',
    'meta_data__og__type', 'meta_data__twitter__description',
    'meta_data__twitter__creator', 'meta_data__twitter__url',
    'meta_data__twitter__image', 'meta_data__twitter__title',
    'meta_data__fb__admins', 'meta_data__twitter__site',
    'meta_data__twitter__card', 'meta_data__robots',
    'meta_data__DC.date.issued', 'meta_data__fb__app_id',
    'meta_data__fb__pages', 'meta_data__article__publisher',
    'meta_data__article__author', 'meta_data__keywords',
    'meta_data__google-site-verification', 'meta_data__article__sectio
    'meta_data__article__tag', 'meta_data__article__published_time',
    'meta_data__article__modified_time', 'meta_data__viewport','Unname
```

In [190]: `fakebf`

Out[190]:

	text	authors__001	canonical_link	
0	Former President Bill Clinton and his Clinton ...	View All Posts	http://eaglerising.com/36899/charity-clinton-f...	Charity: C Found Distri "W
1	After collapsing just before trying to step in...	View All Posts	http://eaglerising.com/36880/a-hillary-clinton...	A Hillary C Administ May be Er
2	Donald Trump is, well, deplorable. He's sugges...	John Prager	http://addictinginfo.com/2016/09/19/trumps-lat...	Trump's Cam Promise M His
3	Website is Down For Maintenance	missing	missing	Website is For Mainte
4	When President Obama spoke at the United Natio...	missing	http://allenwestrepublic.com/obama-pushes-one-...	Obama P One Govern

In [191]:
```python
fakepf=fakepf.drop(columns=[
        'authors__002', 'authors__003', 'authors__004', 'authors__005',
        'authors__006', 'authors__007', 'meta_data__description',
        'meta_data__generator', 'meta_data__author', 'meta_data__og__site_
        'meta_data__og__description', 'meta_data__og__title',
        'meta_data__og__locale', 'meta_data__og__image',
        'meta_data__og__updated_time', 'meta_data__og__url',
        'meta_data__og__type', 'meta_data__twitter__description',
        'meta_data__twitter__creator', 'meta_data__twitter__url',
        'meta_data__twitter__image', 'meta_data__twitter__title',
        'meta_data__fb__admins', 'meta_data__twitter__site',
        'meta_data__twitter__card', 'meta_data__robots',
        'meta_data__DC.date.issued', 'meta_data__fb__app_id',
        'meta_data__fb__pages', 'meta_data__article__publisher',
        'meta_data__article__author', 'meta_data__keywords',
        'meta_data__google-site-verification', 'meta_data__article__sectio
        'meta_data__article__tag', 'meta_data__article__published_time',
        'meta_data__article__modified_time', 'meta_data__viewport','Unname
```

In [192]: `fakepf`

Out[192]:

	text	meta_data__shareaholic__site_name	meta_data__shareaholic__language
0	BREAKING!\n\nLiberal rag Huffington Post is re...	missing	missing
1	Three women who all went missing in the mid-19...	missing	missing
2	On Monday, Bumble Bee Foods and 2 employees we...	missing	missing
3	Republican Rep. Trey Gowdy, who sits on	missing	missing

In [193]: `fakepf.columns`

Out[193]:
```
Index(['text', 'meta_data__shareaholic__site_name',
       'meta_data__shareaholic__language', 'meta_data__shareaholic__url',
       'meta_data__shareaholic__article_author_name',
       'meta_data__shareaholic__image', 'meta_data__shareaholic__site_id',
       'meta_data__shareaholic__article_published_time',
       'meta_data__shareaholic__shareable_page',
       'meta_data__shareaholic__article_modified_time',
       'meta_data__shareaholic__keywords',
       'meta_data__shareaholic__wp_version', 'meta_data__ca_image',
       'meta_data__title', 'authors__001', 'meta_data__outbrainsection',
       'meta_data__msapplication-TileColor', 'meta_data__MobileOptimized',
       'meta_data__googlebot', 'authors__008', 'authors__009', 'authors__010',
       'meta_data__og__image__secure_url', 'meta_data__og__image__width',
       'meta_data__og__image__identifier', 'meta_data__og__image__height',
       'meta_data__twitter__domain', 'meta_data__twitter__text__title',
       'meta_data__twitter__partner', 'meta_data__verifyownership',
       'meta_data__apple-touch-fullscreen',
```

```
       'meta_data__apple-mobile-web-app-capable', 'meta_data__theme-co
lor',
       'meta_data__style-tools', 'meta_data__sailthru.description',
       'meta_data__ca_title', 'meta_data__msapplication-tooltip',
       'meta_data__msapplication-window', 'meta_data__msapplication-ta
sk',
       'meta_data__sailthru.date', 'meta_data__sailthru.title',
       'meta_data__msvalidate.01', 'meta_data__HandheldFriendly',
       'meta_data__specificfeeds-verification-code-ai9qSkdQTFFZc0JCQ0h
CWmw2YVFnL1hPVHQ0aGhWaEVrMTJ2Z3ZwMDNpUFRhbHhWMFdQVmoxK3RPOXNCeEQvdW1Ea
CtWZ2Z2ZHFkWW1jTGM2akhJZkJKClZCaXJxRHJ5VEJNOTdyWDBMYnJkcHVRWFcrbFFQ1RFB
qcmV2aHBHdHB8MVhPS1hCa0JIdjFJbzE3VXNyeDdzZ0syRys3YlcrTjNSdThwWUZUSUhUUY
z0=',
       'meta_data__twitter__image__src', 'meta_data__twitter__image__h
eight',
       'meta_data__twitter__image__width', 'meta_data__propeller',
       'meta_data__ia__markup_url', 'meta_data__sailthru.tags',
       'meta_data__fb__op-recirculation-ads',
       'meta_data__msapplication-TileImage', 'meta_data__onesignal',
       'meta_data__application-name', 'canonical_link', 'images__-', '
title',
       'url', 'movies__-', 'publish_date__$date', 'source', 'fakeness'
],
      dtype='object')
```

```
In [194]: f.drop(columns=['meta_data__shareaholic__site_name',
          _data__shareaholic__language', 'meta_data__shareaholic__url',
          _data__shareaholic__article_author_name',
          _data__shareaholic__image', 'meta_data__shareaholic__site_id',
          _data__shareaholic__article_published_time',
          _data__shareaholic__shareable_page',
          _data__shareaholic__article_modified_time',
          _data__shareaholic__keywords',
          _data__shareaholic__wp_version', 'meta_data__ca_image',
          _data__title','meta_data__outbrainsection',
          _data__msapplication-TileColor', 'meta_data__MobileOptimized',
          _data__googlebot', 'authors__008', 'authors__009', 'authors__010',
          _data__og__image__secure_url', 'meta_data__og__image__width',
          _data__og__image__identifier', 'meta_data__og__image__height',
          _data__twitter__domain', 'meta_data__twitter__text__title',
          _data__twitter__partner', 'meta_data__verifyownership',
          _data__apple-touch-fullscreen',
          _data__apple-mobile-web-app-capable', 'meta_data__theme-color',
          _data__style-tools', 'meta_data__sailthru.description',
          _data__ca_title', 'meta_data__msapplication-tooltip',
          _data__msapplication-window', 'meta_data__msapplication-task',
          _data__sailthru.date', 'meta_data__sailthru.title',
          _data__msvalidate.01', 'meta_data__HandheldFriendly',
          _data__specificfeeds-verification-code-ai9qSkdQTFFZc0JCQ0hCWmw2YVFnL1hPVHQ
          _data__twitter__image__width', 'meta_data__propeller',
          _data__ia__markup_url', 'meta_data__sailthru.tags',
          _data__fb__op-recirculation-ads',
          _data__msapplication-TileImage', 'meta_data__onesignal',
          _data__application-name','images__-','movies__-'])
```

In [195]: `fakepf`

Out[195]:

	text	authors__001		canonical_link
0	BREAKING!\n\nLiberal rag Huffington Post is re...	missing	http://usasnich.com/breaking-putin-interferenc...	INTERFE COU C LEGA
1	Three women who all went missing in the mid-19...	The Potatriot Is A Constitutional Conservative...	http://freedumjunkshun.com/breaking-barrels-re...	Remov Clinton Co
2	On Monday, Bumble Bee Foods and 2 employees we...	missing	missing	Bum Reca Employe
3	Republican Rep. Trey Gowdy, who sits on the Ho...	missing	missing	JUST IN Trey G T DIRECT

In [196]:
```python
# Concat the datasents
df_allr = df_buzzr.append(df_polir, ignore_index=True)
df_allf = fakebf.append(fakepf,ignore_index=True)
df_all=df_allr.append(df_allf,ignore_index=True)
```

In [197]: `df_all.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 437 entries, 0 to 436
Data columns (total 8 columns):
text                  437 non-null object
authors__001          437 non-null object
canonical_link        437 non-null object
title                 437 non-null object
url                   437 non-null object
publish_date__$date   437 non-null object
source                437 non-null object
fakeness              437 non-null int64
dtypes: int64(1), object(7)
memory usage: 27.4+ KB
```

In [198]: `df_all.head(10)`

Out[198]:

	text	authors__001		canonical_link	title
					Donald

0	Less than a day after protests over the police...	More Candace	http://abcnews.go.com/Politics/donald-trump-dr...	Trump: Drugs a 'Very, Very Big Factor' ...
1	Obama To UN: 'Giving Up Liberty, Enhances Secu...	Cassy Fiano	http://rightwingnews.com/barack-obama/obama-un...	Obama To UN: 'Giving Up Liberty, Enhances Secu...	http://rigl
2	Getty Images Wealth Of Nations Trump vs. Clint...	Jack Shafer	http://www.politico.com/magazine/story/2016/09...	Trump vs. Clinton: A Fundamental Clash over Ho...
3	President Obama today vetoed a bill that would...	John Parkinson	http://abcnews.go.com/Politics/president-obama...	President Obama Vetoes 9/11 Victims Bill, Sett...
4	CHAOS! NC Protest MOB Ambushes Female Truck Dr...	Cassy Fiano	http://rightwingnews.com/black-lives-matter/ch...	CHAOS! NC Protest MOB Ambushes Female Truck Dr...	http://ri
5	Hillary Clinton and Donald Trump ushered the 2...	More Veronica	http://abcnews.go.com/Politics/presidential-de...	10 Moments That Mattered From Hillary Clinton ...	h
6	Peaceful protesters crowded Charlotte's first ...	More Michael	http://abcnews.go.com/US/young-girls-emotional...	Young Girl's Emotional Council Speech Laments ...
7	Story highlights Bush will deliver his first l...	Ashley Killough	http://www.cnn.com/2016/09/27/politics/jeb-bus...	Jeb Bush to lecture at Harvard this fall
8	Hillary Clinton's campaign is making one plea ...	Jack Shafer	http://www.politico.com/story/2016/09/clinton-...	Clinton vs. Trump: The debate before the debate
9	I've watched every	Jack Shafer	http://www.politico.com/magazine/story/2016/09...	Is Hillary in a No-Win

presidential
debate
ever br...

Situation?

```
In [199]:  df_all
```

Out[199]:

	text	authors__001	canonical_link	titl
0	Less than a day after protests over the police...	More Candace	http://abcnews.go.com/Politics/donald-trump-dr...	Donald Trump: Drug a 'Very, Ver' Big Factor' .
1	Obama To UN: 'Giving Up Liberty, Enhances Secu...	Cassy Fiano	http://rightwingnews.com/barack-obama/obama-un...	Obama T UN: 'Givin' Up Liberty Enhance Secu.
2	Getty Images Wealth Of Nations Trump vs. Clint...	Jack Shafer	http://www.politico.com/magazine/story/2016/09...	Trump vs Clinton: Fundamenta Clash ove Ho.
3	President Obama today vetoed a bill	John	http://abcnews.go.com/Politics/president-	Presiden Obam Vetoes 9/1

```
In [200]:  df_all.to_csv("/Users/sonia/Downloads/finalyearproject/Complete_DataSet_C
```

```
In [201]:  # Preparing the target and predictors for modeling

           X_body_text = df_all.text.values
           X_headline_text = df_all.title.values
           y = df_all.fakeness.values
```

```
In [220]:  LP-TfidVectorizer to understand the importance of a word in an entire data
           df = TfidfVectorizer(stop_words=ENGLISH_STOP_WORDS,ngram_range=(1,2),max_d
```

```
In [221]:  #
           X_body_tfidf = tfidf.fit_transform(X_body_text)
           X_headline_tfidf = tfidf.fit_transform (X_headline_text)
```

```
In [236]:  X_headline_tfidf_train, X_headline_tfidf_test, y_headline_train, y_headli
           X_body_tfidf_train, X_body_tfidf_test, y_body_train, y_body_test = train_
```

In [237]:
```python
#logistic Regression for headline
lr_headline = LogisticRegression()
```

In [238]:
```python
# train model
lr_headline.fit(X_headline_tfidf_train, y_headline_train)

# get predictions for article section
y_headline_pred = lr_headline.predict(X_headline_tfidf_test)
```

```
/anaconda3/lib/python3.6/site-packages/sklearn/linear_model/logistic.p
y:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
  FutureWarning)
```

In [239]:
```python
print metrics
int ("Logistig Regression F1 and Accuracy Scores : \n")
int ( "F1 score {:.4}%".format( f1_score(y_headline_test, y_headline_pred,
int ( "Accuracy score {:.4}%".format(accuracy_score(y_headline_test, y_hea
```

```
Logistig Regression F1 and Accuracy Scores :

F1 score 64.34%
Accuracy score 64.39%
```

In [240]:
```python
#logistic regression for text
lr_body = LogisticRegression()
```

In [241]:
```python
# train model
lr_body.fit(X_body_tfidf_train, y_body_train)

# get predictions for article section
y_body_pred = lr_body.predict(X_body_tfidf_test)
```

```
/anaconda3/lib/python3.6/site-packages/sklearn/linear_model/logistic.p
y:433: FutureWarning: Default solver will be changed to 'lbfgs' in 0.2
2. Specify a solver to silence this warning.
  FutureWarning)
```

In [242]:
```python
#  metrics
print ("Logistig Regression F1 and Accuracy Scores : \n")
print ( "F1 score {:.4}%".format( f1_score(y_body_test, y_body_pred, aver
print ( "Accuracy score {:.4}%".format(accuracy_score(y_body_test, y_body
```

```
Logistig Regression F1 and Accuracy Scores :

F1 score 85.59%
Accuracy score 85.61%
```

In [ ]:
```
ss-validation,precision,recall. Also twitter user profiles and account det
```