In [129]:
```
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,cross_validate
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 sklearn import datasets, linear_model

from matplotlib import pyplot as plt

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

In [85]:
```
# Reading buzzfeedreal dataset
df_buzzr = pd.read_csv("/Users/sonia/Downloads/finalyearproject/Realnews_
```

In [86]:
```
df_buzzr.columns
```

Out[86]:
```
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 [87]:
```
#Reading the politifact real dataset
df_polir = pd.read_csv("/Users/sonia/Downloads/finalyearproject/Realnews_
```

In [88]:
```
df_polir.columns
```

Out[88]:
```
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 [89]:
```
#reading the buzzfeed fake dataset
fakebf=pd.read_csv("/Users/sonia/Downloads/finalyearproject/FakeNews_Clea
```

In [90]:
```
fakebf.columns
```

Out[90]:
```
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 [91]:
```
#reading the Politifact fake dataset
fakepf=pd.read_csv("/Users/sonia/Downloads/finalyearproject/FakeNews_Clea
```

In [92]:
```
fakepf.columns
```

Out[92]:
```
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 [93]:
```
df_buzzr.head(5)
```

Out[93]:

	Unnamed: 0	top_img	text	authors__001	autho
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	Erich
3	29	http://a.abcnews.com/images/US/AP_Obama_BM_201...	President Obama today vetoed a bill that would...	John Parkinson	Mc
4	33	http://rightwingnews.com/wp-content/uploads/20...	CHAOS! NC Protest MOB Ambushes Female Truck Dr...	Cassy Fiano	

5 rows × 87 columns

In [94]:
```
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 [95]:
```
df_buzzr
```

Out[95]:

	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...	M 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 [96]:
```
df_buzzr=df_buzzr.drop(columns=['Unnamed: 0', 'top_img']) #dropping unnec
```

In [97]:
```
df_buzzr=df_buzzr.drop(columns=['meta_data__news_keywords']) #dropping un
```

In [98]:
```
df_buzzr
```

91	team isn’t laughing at Trump anym...	Edward-Isaac Dovere	http://www.politico.com/story/2016/09/barack-o...	laughing at Trump anymore
92	Story highlights Trump has 45%, Clinton 42% an...	Tal Kopan	http://www.cnn.com/2016/09/19/politics/georgia...	Georgia poll: Donald Trump, Hillary Clinton in...
93	There may be a few women out there who enjoy a...	missing	http://addictinginfo.com/2016/09/19/chelsea-ha...	Chelsea Handler Gets The Last Word After RNC C...
94	Off Message Is Donald Trump qualified to be pr...	Jack Shafer	http://www.politico.com/story/2016/09/is-donal...	Is Donald Trump qualified to be president?

95 rows × 8 columns

In [99]:
```
#dropping unnecessary columns
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 [100]:
```
df_polir
```

Out[100]:

	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 [101]:
```
df_polir.columns
```

Out[101]:
```
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 [102]:
```
#dropping unnecessary columns
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 [103]:
```
df_polir=df_polir.drop(columns=['meta_data__news_keywords'])#dropping unn
```

In [104]:
```
df_polir
```

Out[104]:

	text	authors__001	canonical_link	title
0	Famous dog killed in spot she waited a year fo...	missing	http://rightwingnews.com/top-news/famous-dog-k...	Famous dog killed in spot she waited a year fo..
1	Story highlights The House Oversight panel vot...	Tom Lobianco	http://www.cnn.com/2016/09/22/politics/bryan-p...	House oversight panel votes Clinton IT chief i..
2	We are absolutely heartbroken to hear about th...	Nancy Wells	http://newsbake.com/entertainment-news/music-e...	America Just Tragically Lost A Country Music I..
3	Nine years ago, a driver lost control of	Jack Shafer	http://www.politico.com/magazine/story/2016/09...	Monuments to the Battle for the New

In [105]:
```
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 [106]:
```
fakebf
```

86	Advertisement - story continues below\n\nThe f...	Martin Lioll	http://conservativetribune.com/lester-holt-lie...	Cavut Exposed Holt' Dur
87	Well THAT'S Weird. If the Birther movement is ...	Rich Witmer	http://clashdaily.com/2016/09/dear-cnn-ap-2004...	The AP, In Said You Obama BO
88	\n\nThere's a lot to be discussed about last n...	missing	http://thepoliticalinsider.com/first-president...	People N Something About Hi
89	People Noticed Something Odd About Hillary's O...	Lisa Smith	http://rightwingnews.com/top-news/people-notic...	People N Something About Hi

90 rows × 8 columns

In [107]:
```
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 [108]:
```
fakepf
```

Out[108]:

	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 [109]:
```
fakepf.columns
```

Out[109]:
```
Index(['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__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__MobileOptimiz
ed',
       'meta_data__googlebot', 'authors__008', 'authors__009', 'author
s__010',
       'meta_data__og__image__secure_url', 'meta_data__og__image__widt
h',
       'meta_data__og__image__identifier', 'meta_data__og__image__heig
ht',
       '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
CtWZ2Z2ZHFkWW1jTGM2akhJZkJKclZCaXJxRHJ5VEJNOTdyWDBMYnJkcHVRWFcrbFQ1RFB
qcmV2aHBHdHB8MVhPS1hCa0JIdjFJbzE3VXNyeDdzZ0syRys3YlcrTjNSdThwWUZUSUhUY
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 [110]:
```
fakepf=fakepf.drop(columns=['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','meta_data__outbrainsection',
       'meta_data__msapplication-TileColor', 'meta_data__MobileOptimized'
       'meta_data__googlebot', 'authors__008', 'authors__009', 'authors__
       '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-color
       'meta_data__style-tools', 'meta_data__sailthru.description',
       'meta_data__ca_title', 'meta_data__msapplication-tooltip',
       'meta_data__msapplication-window', 'meta_data__msapplication-task'
       'meta_data__sailthru.date', 'meta_data__sailthru.title',
       'meta_data__msvalidate.01', 'meta_data__HandheldFriendly',
       'meta_data__specificfeeds-verification-code-ai9qSkdQTFFZc0JCQ0hCWm
       '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','images__-','movies__-'])
```

In [111]:
```
fakepf
```

115	Kellyanne Conway, counselor to President Trump...	missing	http://londonwebnews.com/2017/06/01/liberal-wo...	Liberal Becaus "St
116	Police in Vernal Heights, Florida, arrested 3-...	Anthony Brooks	http://themiamigazette.com/cannibals-arrested-...	C Ar Floric Eatin
117	Libtard Democrat Al Franken will resign his se...	Please Enter Your Name Here	http://newsfeedpaper.press/2017/06/15/libtard-...	Libtard Franker To F
118	Chinese Lunar Rover Finds No Evidence Of Ameri...	missing	http://www.antinews.in/chinese-lunar-rover-fin...	Chines Rover F Evic

119 rows × 8 columns

In [112]:
```
# Concat the datasets
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 [113]:
```
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 [114]:
```
df_all.head(10)
```

Out[114]:

	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 [115]:
```
df_all
```

ı/story/2016/09/ted-cruz...	9 times Ted Cruz insulted Donald Trump before ...		http://politi.co/2cSUHmo	1474649129000.0
)16/09/23/politics/critica...	Critical counties: Wake County, NC, could put ...		http://cnn.it/2dnoYpa	missing
/story/2016/09/barack-o...	Obama wears hat, breaking 'Politics 101' rule ...		http://politi.co/2dbESny	1474908937000.0
n/story/2016/09/clinton-...	Clinton: 'The next 50 days will determine the ...		http://politi.co/2cmo9Qd	1474367623000.0

In [116]:
```
df_all.to_csv("/Users/sonia/Downloads/finalyearproject/Complete_DataSet_C
```

In [117]:
```
# 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 [118]:
```
# TfidVectorizer to understand the importance of a word in an entire data
tfidf = TfidfVectorizer(stop_words=ENGLISH_STOP_WORDS,ngram_range=(1,2),m
```

In [119]:
```
#
X_body_tfidf = tfidf.fit_transform(X_body_text)
X_headline_tfidf = tfidf.fit_transform (X_headline_text)
```

In [120]:
```
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 [121]:
```
#logistic Regression for headline/title
lr_headline = LogisticRegression()
```

In [122]:
```
# train model
lr_headline.fit(X_headline_tfidf_train, y_headline_train)

# get predictions for headline 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 [123]:
```
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 [124]:
```
#logistic regression for text
lr_body = LogisticRegression()
```

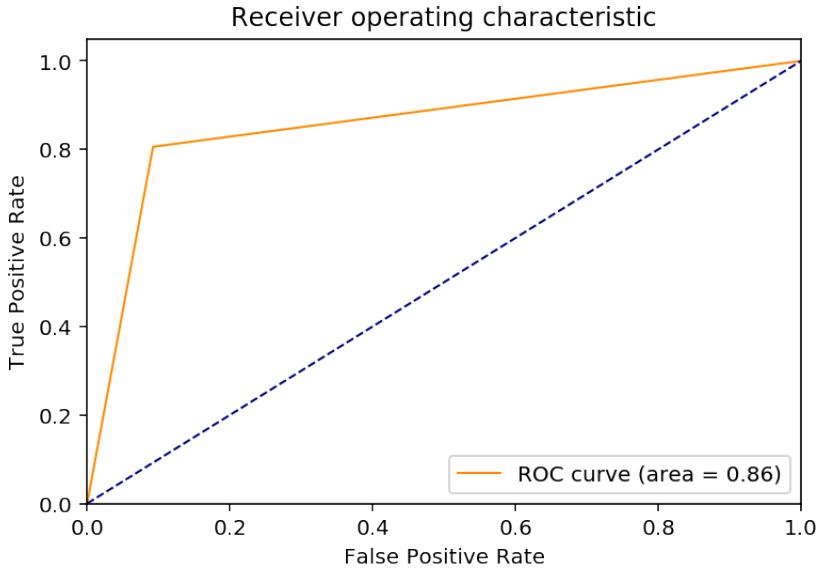
In [125]:
```
# train model
lr_body.fit(X_body_tfidf_train, y_body_train)

# get predictions for text 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 [142]:
```
#  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
prec=precision_score(y_body_test,y_body_pred)
rec=recall_score(y_body_test,y_body_pred)
print(rec)
print(prec)
```

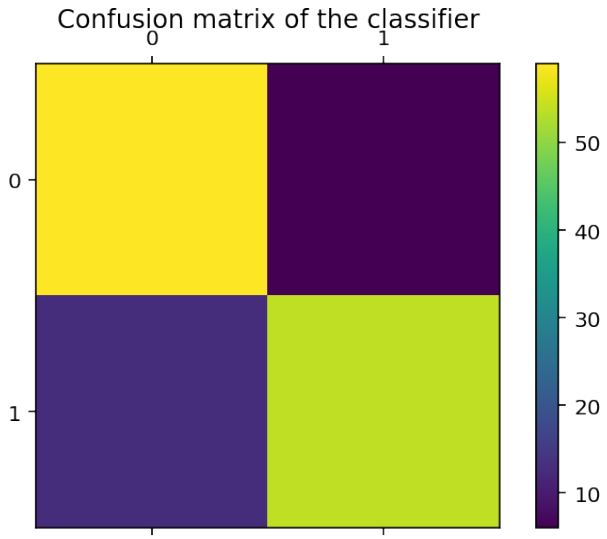
```
Logistig Regression F1 and Accuracy Scores :

F1 score 85.59%
Accuracy score 85.61%
0.8059701492537313
0.9
```

In [137]:
```
from sklearn.metrics import roc_curve, auc
fpr, tpr, thresholds = roc_curve(y_body_test, y_body_pred)
roc_auc = auc(fpr, tpr)

plt.figure()
plt.plot(fpr, tpr, color='darkorange', lw=1, label='ROC curve (area = %0.
plt.plot([0, 1], [0, 1], color='navy', lw=1, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic')
plt.legend(loc="lower right")
plt.show()
```



In [138]:
```
cm = confusion_matrix(y_body_test, y_body_pred)
print(cm)
fig = plt.figure()
ax = fig.add_subplot(111)
cax = ax.matshow(cm)
plt.title('Confusion matrix of the classifier')
fig.colorbar(cax)

plt.show()
```

```
[[59  6]
 [13 54]]
```



In [ ]:
```
urve,confusion matrix,cross-validation,precision,recall. Also twitter user
```

In [78]:
```
#SVM
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn import svm
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.datasets import make_classification

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.text import CountVectorizer
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 sklearn import datasets, linear_model
```

In [29]:
```
df=pd.read_csv("/Users/sonia/Downloads/finalyearproject/Complete_DataSet_
```

In [64]:
```
et `y`
 df.fakeness



rain, X_test, y_train, y_test = train_test_split(df['text'], y, test_size=
```

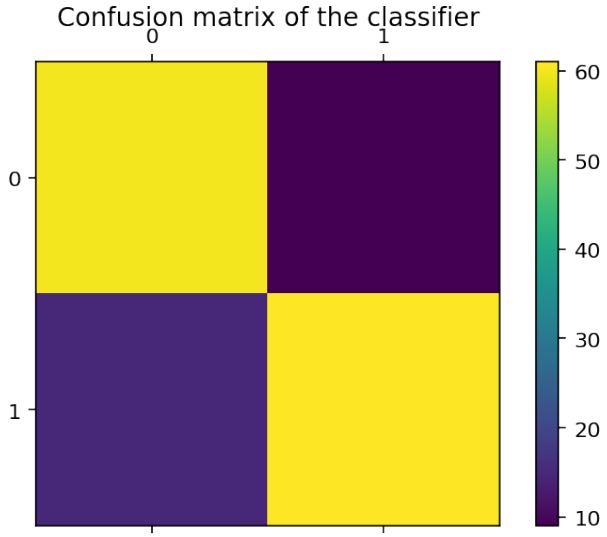
In [65]:
```
# Initialize the `count_vectorizer`
count_vectorizer = CountVectorizer(stop_words='english')

# Fit and transform the training data
count_train = count_vectorizer.fit_transform(X_train)

# Transform the test set
count_test = count_vectorizer.transform(X_test)
```

In [66]:
```
# Initialize the `tfidf_vectorizer`
tfidf_vectorizer = TfidfVectorizer(stop_words='english', max_df=0.7)

# Fit and transform the training data
tfidf_train = tfidf_vectorizer.fit_transform(X_train)

# Transform the test set
tfidf_test = tfidf_vectorizer.transform(X_test)
```

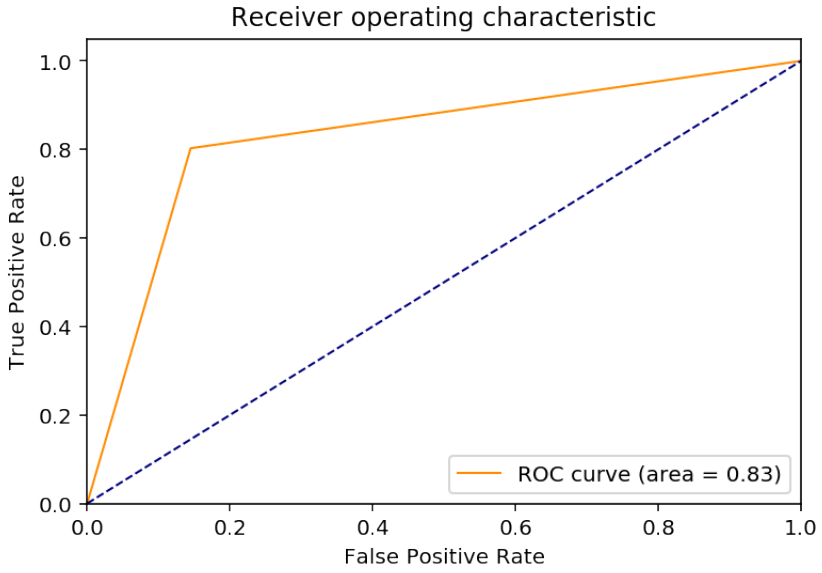
In [141]:
```
linear_clf.fit(tfidf_train, y_train)
pred = linear_clf.predict(tfidf_test)
score = accuracy_score(y_test, pred)
f1score=f1_score(y_test,pred)
prec=precision_score(y_test,pred)
rec=recall_score(y_test,pred)
print("accuracy:   %0.3f" % score)
print("f1_score: %0.3f"% f1score)
print("precision: %0.3f"%prec)
print("rec: %0.3f"%rec)

cm = confusion_matrix(y_test, pred)
print(cm)
fig = plt.figure()
ax = fig.add_subplot(111)
cax = ax.matshow(cm)
plt.title('Confusion matrix of the classifier')
fig.colorbar(cax)

plt.show()
```

```
/anaconda3/lib/python3.6/site-packages/sklearn/linear_model/stochastic
_gradient.py:152: DeprecationWarning: n_iter parameter is deprecated i
n 0.19 and will be removed in 0.21. Use max_iter and tol instead.
  DeprecationWarning)
/anaconda3/lib/python3.6/site-packages/sklearn/linear_model/stochastic
_gradient.py:152: DeprecationWarning: n_iter parameter is deprecated i
n 0.19 and will be removed in 0.21. Use max_iter and tol instead.
```

```
  DeprecationWarning)

accuracy:   0.834
f1_score: 0.836
precision: 0.871
rec: 0.803
[[60  9]
 [15 61]]
```



In [139]:
```
from sklearn.metrics import roc_curve, auc
fpr, tpr, thresholds = roc_curve(y_test, pred)
roc_auc = auc(fpr, tpr)

plt.figure()
plt.plot(fpr, tpr, color='darkorange', lw=1, label='ROC curve (area = %0.
plt.plot([0, 1], [0, 1], color='navy', lw=1, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic')
plt.legend(loc="lower right")
plt.show()
```



In [ ]: