

Fake Profile Detection



TEAM: Sunil Sai Kumar Mada (20MCB1013)
Bhargavinath Dornadula (20MCB1011)
Renuga keerthi (20MCB1017)

Collecting data from Twitter

```
1 import pandas as pd
2 import tweepy
3 import time
4 import numpy as np
5 from tweepy import Stream
6 from tweepy.streaming import StreamListener
7 import numpy as np
8 from sklearn.model_selection import train_test_split
9 from sklearn.utils.multiclass import unique_labels

[ ] 1 consumer_key = 'd9Ksoz6Wb1jD0mqbw8rjaSNb7'
    2 consumer_secret = 'pHXnVSJeLbOxaY1bOR7BWFdDNhZSF6IzegZV87qUSUqy6Qe8qG'
    3 access_token = '3648603434-dGRuInHet22tdoYeqaAGoN8MyZrNw9oXZQvGZUD'
    4 access_token_secret = 'PZ8pcQBCb5zVPLRQNVQZc3Yzi0rz1wPef607RO7gzcvOf'
    5 auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
    6 auth.set_access_token(access_token, access_token_secret)
    7 api = tweepy.API(auth, wait_on_rate_limit=True)

[ ] 1 api

<tweepy.api.API at 0x7f7994afd710>
```

- Collecting data from twitter using tweepy model.
- For that we need Consumer key, Consumer secret key, Access token, Access token.
- We can get those data from twitter app.developer where we have to sign in and make an account and then twitter generate the keys.
- We will import all other packages required

Printing all the friends names of the user

```
1 # printing all the friends names of the user
2 print('Name of the Friends of user')
3 friends = []
4 for friend in tweepy.Cursor(api.friends, screen_name = 'Rajinikanth').items(30):
5     try:
6         friends.append(friend.screen_name)
7         print(friend.screen_name)
8         time.sleep()
9     except Exception as e:
10         pass
11 with open("Fri_1.txt", "w") as f:
12     for item in friends:
13         f.write("%s\n" % item)
```

Output:

```
Name of the Friends of user
anirudhofficial
arrahan
dhanushkraj
soundaryaarajni
ash_r_dhanush
SrBachchan
PMOIndia
narendramodi
firstpost
bsindia
airnewsalerts
EconomicTimes
ABPNews
BBCHindi
aajtak
ZeeNewsEnglish
TimesNow
htTweets
IndiaToday
IndianExpress
ndtv
the_hindu
timesofindia
CNNnews18
```

We try printing the friend names of the particular user. Here, in this case we have considered example of Rajinikanth and have stored the list of names in a file “Fri_1”

1 Data.head()																		
	TwittID	TextData	TweetCreatedAt	RetweetCount	TweetFavouriteCount	TweetSource	UserID	UserScreenName	UserName	UserCreatedAt	UserDescription	UserDescriptionLength	UserFollowersCount	UserFriendsCount	UserLocation	HttpCount	HashtagCount	Menti
0	1399625976800153600	RT @karthiksubbaraj: Here's #JagameThandhiramT...	2021-06-01 07:16:51	3227	0	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Music Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	1	2	
1	1399625927089217536	RT @dhanushk Raja: What would have been a great...	2021-06-01 07:16:39	7527	0	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Music Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	0	0	
2	1395367947552522241	This track and it's video are 🔥 \nAwesome work...	2021-05-20 13:16:58	302	3645	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Music Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	1	1	
3	1395030557092769794	#URGENT \nEmergency need of 72 vials of #Amph...	2021-05-19 14:56:17	925	2715	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Music Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	1	2	
4	1392771872056479748	RT @soundaryaarajni: Our Thalaivar gets his va...	2021-05-13 09:21:05	4565	0	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Music Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	0	2	

TwittID

TextData

TweetCreatedAt

RetweetCount

TweetFavouriteCount

TweetSource

UserID

UserScreenName

UserName

UserCreatedAt

UserDescription

UserDescriptionLength

UserFollowersCount

UserFriendsCount

UserLocation

HttpCount

HashtagCount

MentionCount

TweetCount

Top 5 tweets

Spam profiles detection

AnyDesk AnyDesk NoSqlLip... (2) Aim 2 ... twitter - ... sunil@ka... Telegra... 01:10 PM 27%

Fake-Profile-Detection/T... Fake-Profile-Detection/T... (2) Aim 2 Please Back-Up

twitter.com/aim2pleaseyou2

Apps Tweak And Tric... BlackHatWorld... Google Cyber Security... https://kbr-sjo... Computer Fun... W3Schools OnL... Google Reading list

Home Explore Notifications Messages Bookmarks Lists Profile More

Aim 2 Please Back-Up
33 Tweets

FOLLOW
aim2pleaseyou

Aim 2 Please Back-Up
@aim2pleaseyou2

BACK-UP ACCOUNT FOR @Aim2PleaseYou Just in case I lose my main account... PLEASE FOLLOW

London, UK Joined April 2018

304 Following 1,029 Followers

Not followed by anyone you're following

Tweets Tweets & replies Media Likes

Aim 2 Please Back-Up @aim2pleaseyou2 · Mar 26, 2020
This is my Back-Up account for @Aim2PleaseYou. I would really appreciate

Search Twitter

You might like

- B3n4life @B3n4life1 Follow
- Lingerie Fighting Ch... @LingerieFC Follow
- BBW @BBW22863035 Follow

Show more

What's happening

Technology · Last night
Apple announces new updates and features at Worldwide Developers Conference

Messages

Spam profile-2

Browser tabs: AnyDesk, AnyDesk, NoSqlI..., Fake-P..., twitter ..., (2) Dan..., sunil@..., Telegr..., 01:30 PM, 44%

Browser address bar: twitter.com/Kannibal316

Browser bookmarks: Apps, Tweak And Tric..., BlackHatWorld..., Google, Cyber Security..., https://krb-sjo..., Computer Fun..., W3Schools OnL..., Google, Reading list

Twitter Profile: Daniel N. (@Kannibal316)

Profile picture: A circular profile picture featuring a stylized, colorful, abstract design.

Header image: A large, vibrant, abstract image featuring a character in a red and white helmet, possibly from a video game or anime.

Stats: 2,424 Tweets

Buttons: ... (More options), Follow

Bio: Daniel N. @Kannibal316

Location: Lvl. 35 - Biodegradable

Website: Morrison Hotel, tomarayascream.com

Joined: December 2016

Following: 2,005 | Followers: 2,120

Status: Not followed by anyone you're following

Navigation tabs: Tweets, Tweets & replies, Media, Likes

Recent Tweet: Daniel N. Retweeted Séptimo Game @Septimo_Game - Jun 6

Right sidebar: Search Twitter

Grid of images: A 3x3 grid of various images, including a person, a game, a person, a person, a person, a person, a person, a person, a person.

Section: You might like

Recommendations:

- Nicole Salinas Á. @nicolesalinasa [Follow]
- El Monitor Chile @elmonitorchile [Follow]

Section: What's happening

Messages

Genuine Users:

	TwittID	TextData	TweetCreatedAt	RetweetCount	TweetFavouriteCount	TweetSource	UserID	UserScreenName	UserName	UserCreatedAt	UserDescription	UserDescriptionLength	UserFollowersCount	UserFriendsCount	UserLocation	HttpCount	HashtagCount	Me
0	1399625976800153600	RT @karthiksubbaraj: Here's #JagameThandhiramT...	2021-06-01 07:16:51	3227	0	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	1	2	
1	1399625927089217536	RT @dhanushkraj: What would have been a great...	2021-06-01 07:16:39	7527	0	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	0	0	
2	139536794755252241	This track and it's video are 🔥 inAwesome work...	2021-05-20 13:16:58	302	3645	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Music Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	1	1	
3	1395030557092769794	#URGENT 🚑 inEmergency need of 72 vials of #Amph...	2021-05-19 14:56:17	925	2715	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Music Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	1	2	
4	1392771872056479748	RT @soundaryaarajni: Our Thalaivar gets his va...	2021-05-13 09:21:05	4565	0	Twitter for iPhone	216447259	anirudhofficial	Anirudh Ravichander	2010-11-16 19:13:06	Music Composer/Singer in the Indian Film Industry	49	8138319	207	Chennai	0	2	
...
715	1402144157053853703	RT @news18dotcom: @AmankayamHal_ In the detail...	2021-06-08 06:03:12	2	0	TweetDeck	6509832	CNNnews18	News18	2007-06-01 20:31:01	Lightning fast alerts, #BreakingNews from Indi...	62	4597047	403	India	0	0	
716	1402143957325258753	RT @news18dotcom: @AmankayamHal_ News18's @Ama...	2021-06-08 06:02:24	3	0	TweetDeck	6509832	CNNnews18	News18	2007-06-01 20:31:01	Lightning fast alerts, #BreakingNews from Indi...	62	4597047	403	India	0	0	
717	1402142378102640643	It is learnt that the Centre had planned to am...	2021-06-08 05:56:08	3	11	Twitter Web App	6509832	CNNnews18	News18	2007-06-01 20:31:01	Lightning fast alerts, #BreakingNews from Indi...	62	4597047	403	India	1	0	
718	1402140713614299136	It was propagated as Modi vaccine: @RajeevRai...	2021-06-08 05:49:31	1	7	Grabyo	6509832	CNNnews18	News18	2007-06-01 20:31:01	Lightning fast alerts, #BreakingNews from Indi...	62	4597047	403	India	1	0	
719	1402140137312587777	RT @news18dotcom: An immunocompromised individ...	2021-06-08 05:47:13	1	0	TweetDeck	6509832	CNNnews18	News18	2007-06-01 20:31:01	Lightning fast alerts, #BreakingNews from Indi...	62	4597047	403	India	1	0	

Count of spam users and genuine users:

Total count : 930

```
1 df.shape  
(930, 25)
```

Genuine : 720

```
1 Total_leg_data.shape  
(720, 24)
```

Fake: 210

```
1 Total_spam_data.shape  
(210, 25)
```


Attributes that we have in our dataset now are:

```
[ ] 1 df.columns
```

```
Index(['TwittID', 'TextData', 'TweetCreatedAt', 'RetweetCount',  
      'TweetFavouriteCount', 'TweetSource', 'UserID', 'UserScreenName',  
      'UserName', 'UserCreatedAt', 'UserDescription', 'UserDescriptionLength',  
      'UserFollowersCount', 'UserFriendsCount', 'UserLocation', 'HttpCount',  
      'HashtagCount', 'MentionCount', 'TweetCount', 'AvgHashtag',  
      'AvgURLCount', 'AvgMention', 'AvgRetweet', 'AvgFavCount',  
      'SpammerOrNot'],  
      dtype='object')
```

Attributes that are been considered for training model

```
[ ] 1 Lis = ['RetweetCount','TweetFavouriteCount','UserDescriptionLength','UserFollowersCount','UserFriendsCount','HttpCount','HashtagCount','MentionCount','TweetCount','AvgHashtag','AvgURLCount','AvgMention','AvgRetweet','AvgFavCount']
```

```
[ ] 1 X = df[Lis]
```

```
[ ] 1 X
```

	RetweetCount	TweetFavouriteCount	UserDescriptionLength	UserFollowersCount	UserFriendsCount	HttpCount	HashtagCount	MentionCount	TweetCount	AvgHashtag	AvgURLCount	AvgMention	AvgRetweet	AvgFavCount
0	3227.0	0	49	8138319	207	1	2	4	12891	1.133333	0.433333	1.566667	2806.4	6632.266667
1	7527.0	0	49	8138319	207	0	0	1	12891	1.133333	0.433333	1.566667	2806.4	6632.266667
2	302.0	3645	49	8138319	207	1	1	3	12891	1.133333	0.433333	1.566667	2806.4	6632.266667
3	925.0	2715	49	8138319	207	1	2	0	12891	1.133333	0.433333	1.566667	2806.4	6632.266667
4	4565.0	0	49	8138319	207	0	2	1	12891	1.133333	0.433333	1.566667	2806.4	6632.266667
...
925	0.0	0	98	275	922	1	0	3	8481	0.166667	0.600000	1.133333	65.9	0.400000
926	0.0	5	98	275	922	1	0	1	8481	0.166667	0.600000	1.133333	65.9	0.400000
927	25.0	0	98	275	922	1	0	1	8481	0.166667	0.600000	1.133333	65.9	0.400000
928	0.0	3	98	275	922	0	0	3	8481	0.166667	0.600000	1.133333	65.9	0.400000
929	17.0	0	98	275	922	1	0	1	8481	0.166667	0.600000	1.133333	65.9	0.400000

930 rows × 14 columns

ML model :

Logistic Regression

```
1 X_train,X_test,y_train,y_test = train_test_split(X, y, test_size=0.20, random_state=24)

1 from sklearn.linear_model import LogisticRegression
2

1 model=LogisticRegression(max_iter=1000)

1 model.fit(X_train, y_train)

LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                    intercept_scaling=1, l1_ratio=None, max_iter=1000,
                    multi_class='auto', n_jobs=None, penalty='l2',
                    random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                    warm_start=False)

1 prediction=model.predict(X_test)

1 model

LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                    intercept_scaling=1, l1_ratio=None, max_iter=1000,
                    multi_class='auto', n_jobs=None, penalty='l2',
                    random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                    warm_start=False)
```

Accuracy:

We have used logistic regression for prediction and got accuracy of 95.33

```
[166] prediction
```

```
array([0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1,
       0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1,
       0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1,
       1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0])
```

```
[170] Acc = accuracy_score(y_test, prediction)
```

```
[173] print("LR Model Acc : ",Acc*100)
```

```
LR Model Acc : 95.33
```

Thank you