anfederico / Stocktalk

Data collection toolkit for social media analytics

#data-mining #sentiment-analysis #twitter

231 commits | 1 branch | 0 releases | 1 contributor | MIT

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

anfederico Update README.md — Latest commit 1c2c585 25 days ago

docs	Demo of visualization	26 days ago
media	Delete Demo.gif	26 days ago
stocktalk	Update settings.py	26 days ago
LICENSE.txt	Update LICENSE.txt	26 days ago
README.md	Update README.md	25 days ago
requirements.txt	Add files via upload	26 days ago

README.md

Purpose

Stocktalk is a visualization tool that tracks tweet volume and sentiment on Twitter, given a series of queries.

It does this by opening a local websocket with Twitter and pulling tweets that contain user-specified keywords. For example, I can tell Stocktalk to grab all tweets that mention Ethereum and periodically tally volume and measure average sentiment every 15 minutes.

It will then record this data continuously and update an online database that can be used to visualize the timeseries data via an interactive Flask-based web application.

Demo

https://anfederico.github.io/Stocktalk/

Prerequisites

> Stocktalk requires API credentials with Twitter and Mlab

Twitter Steps (Creating an application)

1. Sign into Twitter at apps.twitter.com
2. Create a new application and fill out details
3. Generate an access token
4. Save the following information
    - Consumer Key
    - Consumer Secret
    - Access Token
    - Access Token Secret

Mlab Steps (Setting up an online database)

1. Make an account at https://mlab.com
2. Create a new deployment in sandbox mode
3. Add a database user to your deployment
4. Save the following information
    - Mongo deployment server
    - Mongo deployment id
    - Mongo deployment client
    - Deployment user
    - Deployment pass

Download

```
# Clone repository and install dependencies
$ git clone https://github.com/anfederico/Stocktalk
$ pip install -r Stocktalk/requirements.txt

# Install natural language toolkit sentiment corpus
$ python -m nltk.downloader vader_lexicon
```

Edit Settings

```
/stocktalk
└── /scripts
    └── settings.py
```

```
# Mongo
mongo_server = 'ds254236.mlab.com'
mongo_id     =  54236
mongo_client = 'stocktalk'
mongo_user   = 'username'
```

```
mongo_pass   = 'password'

# Twitter
api_key             = ''
api_secret          = ''
access_token        = ''
access_token_secret = ''
credentials = [api_key, api_secret, access_token, access_token_secret]
```

Code Examples

Twitter Streaming

> This file opens the websocket and writes to the online databse until manually interrupted

```
/stocktalk
└── listen.py

$ python listen.py
```

```
from scripts import settings

# Each key or category corresponds to an array of keywords used to pull tweets
queries = {'ETH': ['ETH', 'Ethereum'],
           'LTC': ['LTC', 'Litecoin'],
           'BTC': ['BTC', 'Bitcoin'],
           'XRP': ['XRP', 'Ripple'],
           'XLM': ['XLM', 'Stellar']}

# Aggregate volume and sentiment every 15 minutes
refresh = 15*60

streaming.streamer(settings.credentials,
                   queries,
                   refresh,
```

```
                    sentiment=True,
                    debug=True)
```

Realtime Visualization

> This file initiates a local web-application which pulls data from the online database

```
/stocktalk
└── app.py

$ python app.py
```

Underlying Features

Text Processing

```
t1 = "@TeslaMotors shares jump as shipments more than double! #winning"
print(process(t1))

t2 = "Tesla announces its best sales quarter: http://trib.al/RbTxvSu $TSLA"
print(process(t2))

t3 = "Tesla $TSLA reports deliveries of 24500, above most views."
print(process(t3))
```

```
shares jump as shipments more than double winning
tesla announces its best sales quarter
tesla reports deliveries of number above most views
```

Sentiment Analysis

```
t1 = "shares jump as shipments more than double winning"
print(sentiment(t1))

t2 = "tesla reports deliveries of number above most views"
print(sentiment(t2))

t3 = "not looking good for tesla competition on the rise"
print(sentiment(t3))
```

```
0.706
0.077
-0.341
```