Crypto sentiment Analysis

On Twitter

VORAUSSETZUNGEN

- Erstellen eines Twitter Dev Accounts

- Erstellen einer Twitter app
- Speichern der Private und public keys für die API Verbindung
- Aufrufen und generieren der Authentication tokens

Hier in einem Ubuntu Linux OS

- Python3 Installation

```
apt install python3
```

- pip Installation

```
apt install python3-pip
```

- Tweepy Installation

```
pip install tweepy
```

- Kafka Installation

```
pip install kafka-python
```

- Python Twitter installation

```
pip install python-twitter
```

DATENERHEBUNG

<u>Aktuelle/Streamed Daten von Twitter API beziehen</u>

⟶ Voraussetzung Twitter Dev account

1. Hier mit setup einer VM/kann auch lokal bezogen werden
2. Installation von apache Kafka auf Linux OS (beinhaltet Zookeeper)
3. Starten des Kafka Servers

```
~$ <'kafka directory'>/bin/kafka-server-start.sh config/server.properties
```

4. Starten einer zweiten shell um ein topic in Kafka zu erstellen

```
~$ <'kafka directory'>/bin/kafka-topics.sh --create --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1 --topic <'name'>
```

5. Anpassen der 'server.properties'-file (+Bootstrap connection)
6. Schreiben des Python scripts

SCRIPT ZUR DATENERHEBUNG

```
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
from tweepy import Stream
from kafka import KafkaProducer
import json #for easier data manipulation in the end

#Just for authorizsation
access_token = "904485960-gFiS63STA32GpCd7RuEkEeP8RSTd1ZJzvRXWu6Gv"
access_token_secret = "                moQYs5d36dy2eSWUk"

# tweepy library to authenticate our API keys
api_key = "0z5aR2Q5enMDxxhfp6Q0Nj5tI"
api_secret = "                                  imB5H"
```

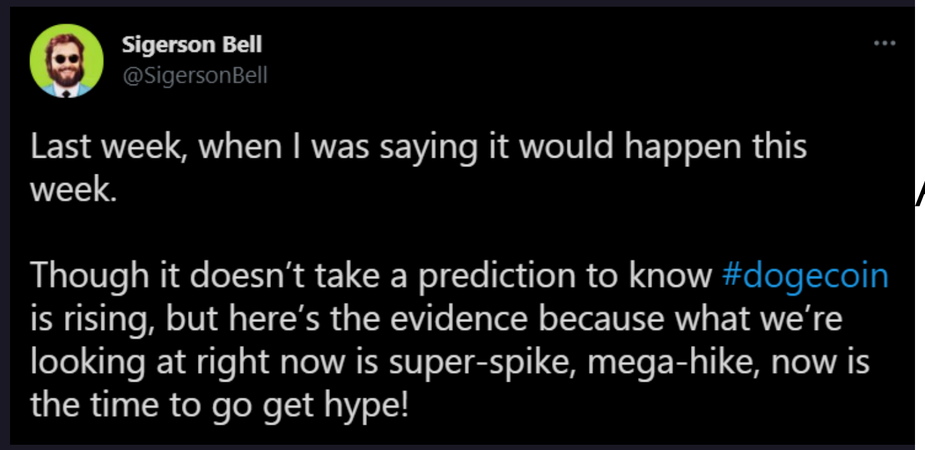
```
#function in class that 'listens' for the kafka topic and writes data in json format to it
class StdOutListener(StreamListener):
    def on_data(self, data):
        json_ = json.loads(data)
        producer.send("doge", data.encode('utf-8'))
        return True

    def on_error(self, status):
        print (status)

#listens to ip which is given in the server.properties file as well as here in link to the kafak topic
producer = KafkaProducer(bootstrap_servers='          :9092')
l = StdOutListener()

#twitter authorisation keys
auth = OAuthHandler(api_key, api_secret)
auth.set_access_token(access_token, access_token_secret)

#Data that should be scanned for in streamed twitter feeds
stream = Stream(auth, l)
stream.filter(track=["doge-coin", "doge", "doge-crypto"])
```

VORLIEGENDE DATEN


```
{
  "created_at":"Mon Jun 28 16:25:36 +0000 2021",
  "id":1409548546102022150,
  "id_str":"1409548546102022150",
  "text":"RT @Lilmook4real: Yo #DogeFather @elonmusk the #DogeArmy and I would like to wish you a happy birthday \ud83c\udf89\ud83c\udf88 I hope this song makes you smile\u2026",
  "source":"\u003ca href=\"https:\/\/mobile.twitter.com\" rel=\"nofollow\"\u003eTwitter Web App\u003c\/a\u003e",
  "truncated":false,
  "in_reply_to_status_id":null,
  "in_reply_to_status_id_str":null,
  "in_reply_to_user_id":null,
  "in_reply_to_user_id_str":null,
  "in_reply_to_screen_name":null,
  "user":{
    "id":1386253909077807105,
    "id_str":"1386253909077807105",
    "name":"I just want a yogurt \ud83e\udda7",
    "screen_name":"justwantayogurt",
    "location":"Reddit Island",
    "url":"https:\/\/www.reddit.com\/user\/ijustwantayogurt",
    "description":"I'm a retard ape who loves & HODL cryptos & stonks. I'm not a Financial Advisor",
    "translator_type":"none",
    "protected":false,
    "verified":false,
    "followers_count":74,
    "friends_count":335,
    "listed_count":0,
    "favourites_count":1231,
    "statuses_count":649,
    "created_at":"Sun Apr 25 09:41:07 +0000 2021",
    "utc_offset":null,
    "time_zone":null,
    "geo_enabled":false,
    "lang":null,
    "contributors_enabled":false,
    "is_translator":false,
    "profile_background_color":"F5F8FA",
    "profile_background_image_url":"",
    "profile_background_image_url_https":"",
    "profile_background_tile":false,
    "profile_link_color":"1DA1F2",
    "profile_sidebar_border_color":"C0DEED",
    "profile_sidebar_fill_color":"DDEEF6",
    "profile_text_color":"333333",
    "profile_use_background_image":true,
    "profile_image_url":"http:\/\/pbs.twimg.com\/profile_images\/1403290160515616771\/FUlLEWut_normal.jpg",
    "profile_image_url_https":"https:\/\/pbs.twimg.com\/profile_images\/1403290160515616771\/FUlLEWut_normal.jpg",
    "profile_banner_url":"https:\/\/pbs.twimg.com\/profile_banners\/1386253909077807105\/1624079754",
    "default_profile":true,
    "default_profile_image":false,
    "following":null,
    "follow_request_sent":null,
    "notifications":null,
    "withheld_in_countries":[
```

BEISPIEL TWEEETS (POSITIV)

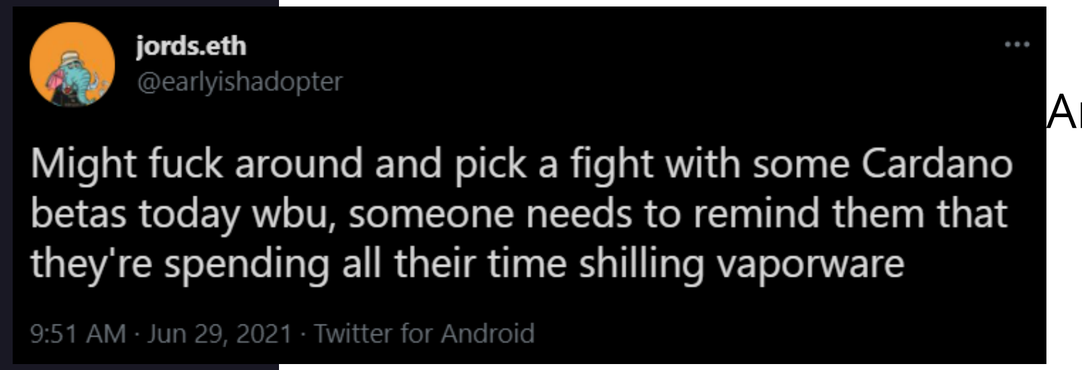
Analysed sentiment: positive



Analysed sentiment: positive

Analysed sentiment: positive

BEISPIEL TWEEETS (NEGATIV)



Analysed sentiment: negativ

Analysed sentiment: negativ

Analysed sentiment: negativ

HUMAN VS. ALGORITHM

Model evaluation: neutral

Human evaluation: eher positiv

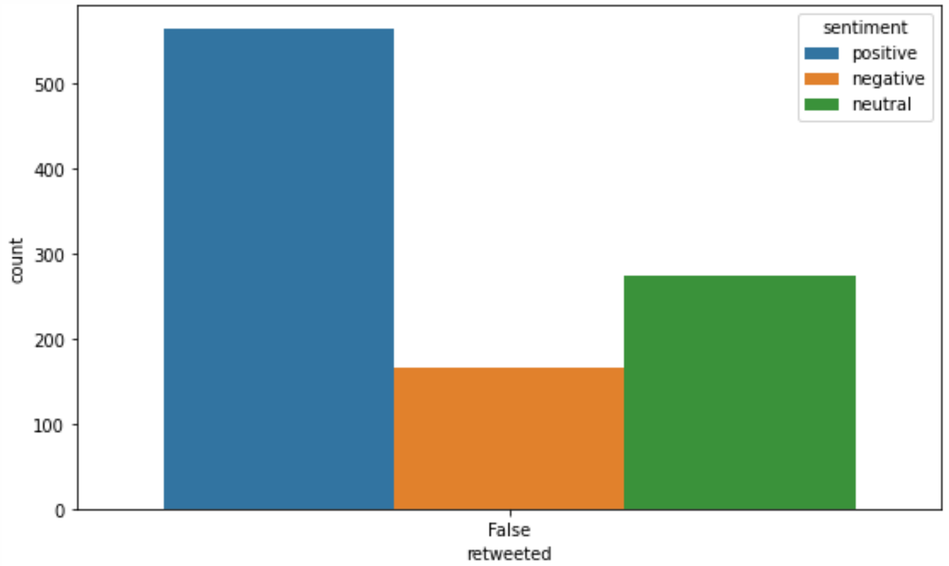
Model evaluation: positiv/neutral

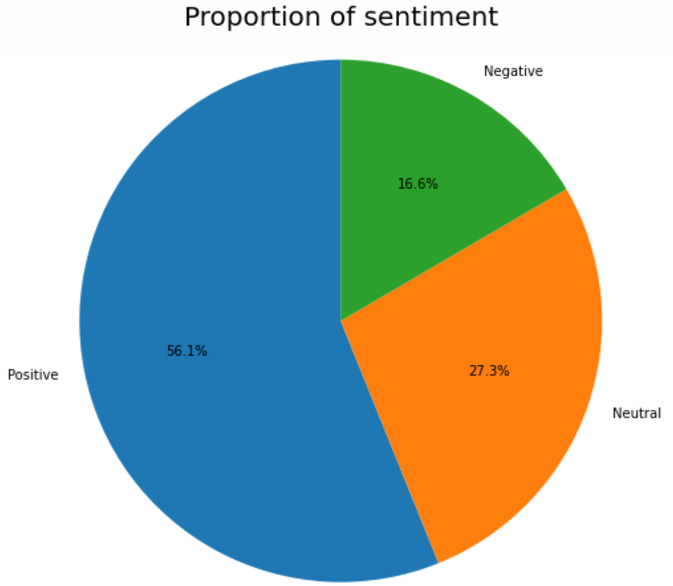
Human evaluation: negativ

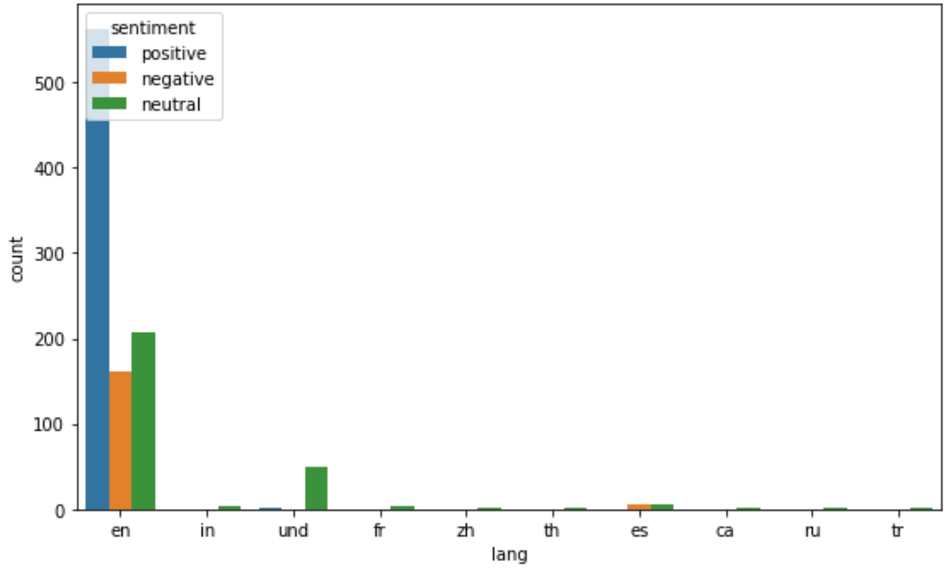
Model evaluation: positiv

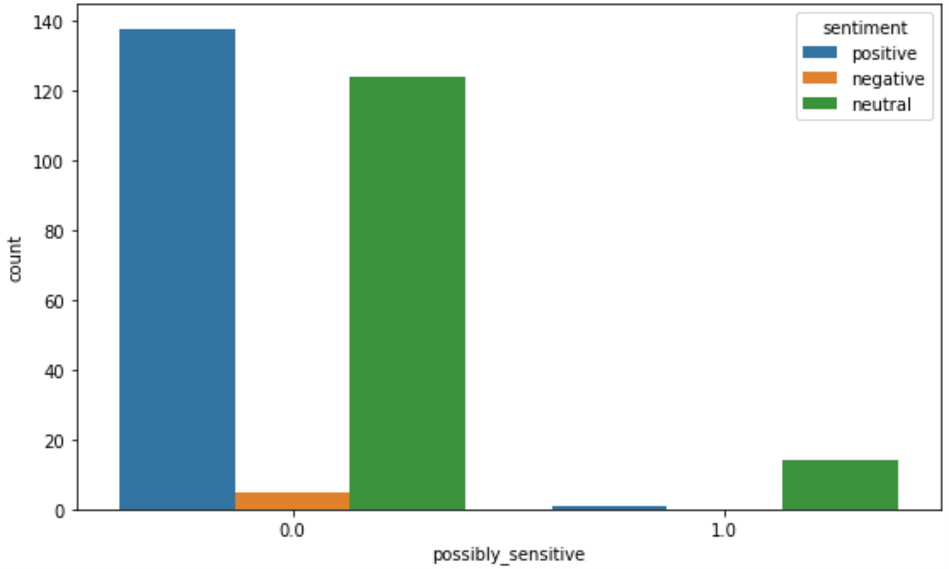
Human evaluation: negativ

DATENAUSWERTUNG









PROBLEME EINES TWITTER SENTIMENTS

- Die Daten sind mehrheitlich von einer Quelle
- Auf Twitter werden hauptsächlich Informationen von Befürwortern veröffentlicht (biased)
- Werbung ist oft schwer zu filtern
- Memes oder Sarkasmus lassen sich von einem Algorithmus nur schwer analysieren
- Ohne das evaluieren durch einen Menschen können schlecht Handlungsempfehlungen geben werden
-

Thank you for your attention!