Sentiment Analyis and Prediction

Index:

Sources

As the grade isn't abouyt how we elaborate the dataset we used a DataSet (https://www.kaggle.com/maxjon/complete-tweet-sentiment-extraction-data) originally made for a competition Thanks to the author of this dataset To interact with the data we will use numpy (https://numpy.org/), pandas (https://pandas.pydata.org/), nltk (https://www.nltk.org/), pickle (https://docs.python.org/3/library/pickle.html) and sklearn (https://scikit-learn.org/stable/) We also used tweepy (https://www.tweepy.org/) to try our trained model

PreProcessing

Setup

In [29]:
```
import numpy as np
import pandas as pd
import sklearn as sk
import nltk
import re

## Tokenization
nltk.download('punkt')

# Normalization
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')

#Cleaning
nltk.download('stopwords')
```

```
[nltk_data] Downloading package punkt to /home/romain/nltk_dat
a...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to /home/romain/nltk_dat
a...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /home/romain/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to
-
[nltk_data]       date!
[nltk_data] Downloading package stopwords to /home/romain/nltk_da
ta...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[29]:
```
True
```

Importation

We import the train data in order to preprocess it

In [30]:
```
train = pd.read_csv("data/train.csv")
# Remove null data
train = train.dropna()

print(train.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 27480 entries, 0 to 27480
Data columns (total 4 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   textID         27480 non-null  object
 1   text           27480 non-null  object
 2   selected_text  27480 non-null  object
 3   sentiment      27480 non-null  object
dtypes: object(4)
memory usage: 1.0+ MB
None
```

Tokenization

In order analyze the string data we 'tokenize' it It means that we split every sentence into an array of word and ponctuation

In [31]:
```
train["text"] = [nltk.tokenize.word_tokenize(i) for i in train["t
ext"]]
print(train.head())
```

```
       textID                                               text
\
0  cb774db0d1  [I, `, d, have, responded, ,, if, I, were, going]
1  549e992a42  [Sooo, SAD, I, will, miss, you, here, in, San,...
2  088c60f138                  [my, boss, is, bullying, me, ...]
3  9642c003ef             [what, interview, !, leave, me, alone]
4  358bd9e861  [Sons, of, *, *, *, *, ,, why, couldn, `, t, t...

                         selected_text sentiment
0  I`d have responded, if I were going   neutral
1                            Sooo SAD  negative
2                          bullying me  negative
3                       leave me alone  negative
4                        Sons of ****,  negative
```

Remove Noise from our dataSet

The noises are things like empty word, spaces, single letter or special characters. What we'll need to do now is to remove all the unwanted data with RegEx for example We may want to lower the words too

In [32]:
```
def cleanData(word : str):
    return re.sub(r'[^A-Za-z0-9_]','',word).lower()
```

Normalization

We now have all our sentence sliced up into words, but we do face a problem. The same word can have multiple forms depending on the context. Our goal will be to transform all those version into the radical in order to get a smaller dictionnary. The process of grouping together forms of a word is called **Lemmatisation**. Firstly we tag the words to identify their type and secondly we use a dictionary to transform them to a simplest form.

StopWords

What we want to do at the same time is removing useless words. Some words might not be helpful for us to understand the whole sentence, those are called stopwords For example: "the", "an", "in" doesn't help and would weaken the model as they appear to be really common in all type of sentence. They doesn't help us taking a decision

In [33]:
```
tag_map = dict()
tag_map['V'] = nltk.corpus.wordnet.VERB
tag_map['J'] = nltk.corpus.wordnet.ADJ
tag_map['N'] = nltk.corpus.wordnet.NOUN
tag_map['R'] = nltk.corpus.wordnet.ADV
tag_map['S'] = nltk.corpus.wordnet.ADJ_SAT

# map the given posTag to the matching postag for nltk
def convertPostagToLemmitizationTag(pos_tag):
    return tag_map.get(pos_tag, nltk.corpus.wordnet.NOUN)
```

In [34]:
```
def lemmatize_sentence(sent : str):
    lemmatizer = nltk.stem.wordnet.WordNetLemmatizer()
    lemmatized_sentence = []
    for word, tag in nltk.pos_tag(sent):
        ## Clean our words with the function we made earlier
        cleanedWord = cleanData(lemmatizer.lemmatize(word, conver
tPostagToLemmitizationTag(tag[0])))
        ## If the word isn't null or a stopword then we add it to
our final array
        if cleanedWord is not None and len(cleanedWord) > 0 and c
leanedWord not in nltk.corpus.stopwords.words('english'):
            lemmatized_sentence.append(cleanedWord)
    return lemmatized_sentence
```

In [35]:
```
## lemmatize and clean the sentence
train["text"] = [lemmatize_sentence(i) for i in train["text"]]
print(train.head())
```

```
       textID                                  text  \
0  cb774db0d1                        [respond, go]
1  549e992a42      [sooo, sad, miss, san, diego]
2  088c60f138                         [bos, bully]
3  9642c003ef           [interview, leave, alone]
4  358bd9e861  [sons, put, release, already, buy]

                          selected_text sentiment
0  I`d have responded, if I were going   neutral
1                            Sooo SAD  negative
2                          bullying me  negative
3                       leave me alone  negative
4                        Sons of ****,  negative
```

Now to get a quick idea of the words repartition, we iterate through all the dataset, take the most used words for each sentiment We do have a pretty good idea of what type of words are mostly used

In [36]:
```
commonDic = dict()
dataSet = {}

for i in ["neutral", "positive", "negative"]:
    commonDic[i] = []
    for sentence in train[train["sentiment"] == i]["text"]:
        commonDic[i] = np.concatenate((commonDic[i], sentence))

    commonDic[i] = nltk.FreqDist(commonDic[i])
    print(i + ": " + str(commonDic.get(i).most_common(10)))
```

```
neutral: [('get', 1264), ('go', 1166), ('day', 645), ('work', 64
0), ('http', 596), ('lol', 494), ('u', 479), ('like', 474), ('tim
e', 471), ('know', 456)]
positive: [('day', 1345), ('good', 1191), ('love', 1012), ('happy
', 866), ('get', 774), ('go', 636), ('thanks', 559), ('mother', 5
33), ('great', 495), ('like', 471)]
negative: [('get', 922), ('go', 836), ('miss', 636), ('work', 49
4), ('like', 490), ('day', 420), ('feel', 408), ('sad', 405), ('i
m', 370), ('bad', 369)]
```

Encoding

To process the data easier we'll use an Encoder, firstly we'll encode the Y data (the sentiment). In order to do that we'll use the sk LabelEncoder

In a second time we do the same for our our words by creating a dictionary of all our worlds and encoding it After that we encode each row But that time we will not use the LabelEncoder, instead we use the CountVectorizer Basically it create a dictionnary of known words, a matrix where each word is a row in the matrix. This way it's way easier to process data in our futur model

In [37]:
```
sentimentEncoder = sk.preprocessing.LabelEncoder()

encodedSentiment = sentimentEncoder.fit_transform(train["sentimen
t"])
```

In [38]:
```
textEncoder = sk.feature_extraction.text.CountVectorizer()
encodedText = textEncoder.fit_transform([ ' '.join(i) for i in tr
ain["text"]])
```

Model

Now that our data is cleaned up, we want to run our Model. As we want to put a label on a given sentence we conclude that we'll be using a classification algortihm.

But Firstly we'll split up the data we previously preprocessed in order to get a training and testing set

In [39]:
```
Xtrain, Xtest, Ytrain, Ytest = sk.model_selection.train_test_spli
t(encodedText, encodedSentiment, test_size=0.15)
```

To find the best matching algorithms, we decide to test out a few classificaiton algorithms on a lightweight dataset

In [12]:
```
from sklearn.naive_bayes import GaussianNB
from sklearn.naive_bayes import MultinomialNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn import svm
from sklearn.neighbors import KNeighborsClassifier
```

In [32]:
```
for model in [GaussianNB(), MultinomialNB(), DecisionTreeClassifi
er(), RandomForestClassifier(), svm.SVC(), KNeighborsClassifier
()]:
    model.fit(Xtrain.toarray()[:1000],Ytrain[:1000])
    prediction = model.predict(Xtest.toarray()[:1000])
    accuracy = sk.metrics.accuracy_score(Ytest[:1000], predictio
n)
    with open(type(model).__name__ + ".result", "w") as f:
        f.write("\n\n" + type(model).__name__ + ": " + str(accura
cy))
        f.write("\nPrediction:\n")
        f.write(str(prediction))
        f.write("\nReal:\n")
        f.write(str(Ytest[:1000]))
    print(type(model).__name__ + ": " + str(accuracy))
```

```
GaussianNB: 0.397
MultinomialNB: 0.56
DecisionTreeClassifier: 0.608
RandomForestClassifier: 0.611
SVC: 0.585
KNeighborsClassifier: 0.473
```

We now have a clearer idea of which type of algorithm we want to use. In order to know which one will best fit our data we decide to train the Top3 with the full set of data.

In [33]:
```
for model in [MultinomialNB(), DecisionTreeClassifier(), RandomFo
restClassifier()]:
    model.fit(Xtrain.toarray(),Ytrain)
    prediction = model.predict(Xtest.toarray())
    accuracy = sk.metrics.accuracy_score(Ytest, prediction)
    with open(type(model).__name__ + ".bigResult", "w") as f:
        f.write("\n\n" + type(model).__name__ + ": " + str(accura
cy))
    print(type(model).__name__ + ": " + str(accuracy))
```

```
MultinomialNB: 0.6491994177583698
DecisionTreeClassifier: 0.6669092673459486
RandomForestClassifier: 0.6933527413876759
```

We now have to decide which algorithm to use in our definitive version, in our case we will use the RandomForestClassifier. By the way, this algorithm is well known to work perfectly with text analysis so it's pretty logic.

Reusability

As we don't want to train our model each time, we train a model and save it with pickle. That way we can load it quickly and make predictions without retraining the model every time.

In [13]:
```
import pickle

model = RandomForestClassifier()
model.fit(Xtrain.toarray(), Ytrain)
prediction = model.predict(Xtest.toarray())
accuracy = sk.metrics.accuracy_score(Ytest, prediction)
with open(type(model).__name__ + ".finalResult", "w") as f:
    f.write("\n\n" + type(model).__name__ + ": " + str(accuracy))
print(type(model).__name__ + ": " + str(accuracy))

pickle.dump(model, open('trainedModel.sav', 'wb'))
```

```
RandomForestClassifier: 0.6926249393498302
```

We now have a pre-trained model that we can use on the run like so:

In [13]:
```
model = pickle.load(open('trainedModel.sav', 'rb'))
result = model.score(Xtest.toarray(), Ytest)
print(result)
```

```
0.9539058709364386
```

Results

In [41]:
```
import pickle

model = pickle.load(open('trainedModel.sav', 'rb'))
result = model.score(Xtest.toarray(), Ytest)
print(result)
```

```
0.9539058709364386
```

As we can see there, our model, once trained is very efficient and as an accuracy score of 95% which is pretty good We are satisfied with this accuracy and keep our train model like so

Conclusion

To conclude, we do have now a model trained to identify the sentiment of a tweet. The most important part is how we processed the data to make a more precise and efficient model. We had to make a few decisions about how we process the data, removing stopwords or not, keeping special characters, encoding algorithm The model we decided to use is also really important, and we do think that the RandomForestClassifier fits our need.

Usage

We can use this tweet sentiment recognizer for multiple usages if combined with a tweet scraper For example, it could be interesting to determine the whole sentiment of a twitter account or by month to know the mindset of a given person But most scrapers are done lately so we'll have to use the official API

In [42]:
```
def get_sentiment(tweet : str):
    encoded_text = textEncoder.transform([' '.join(lemmatize_sent
ence(nltk.tokenize.word_tokenize(tweet)))]) #([ ' '.join(i) for i
in lemmatize_sentence(tweet)])
    return sentimentEncoder.inverse_transform(model.predict(encod
ed_text.toarray()))[0]

print(get_sentiment("i'm really sad those day"))
print(get_sentiment("do french fries have a soul ?"))
print(get_sentiment("so happy to see my friend today"))
```

```
negative
neutral
positive
```

Bonus

Here is a sample code to interact with a user profile to get his tweets positivity score

In [63]:
```
import tweepy
def get_profile_positivity(username : str, nb_tweet : int = 100,
detailled : bool = False):
    auth = tweepy.OAuthHandler(
        open("./info/api.key","r").read(),
        open("./info/apikey.secret","r").read()
    )
    auth.set_access_token(
        open("./info/access.token","r").read(),
        open("./info/accesssecret.token","r").read()
    )

    api = tweepy.API(auth)
    positivity = {
        "negative":0,
        "neutral":0,
        "positive":0
    }
    if detailled:
        positivity["negative"] = []
        positivity["positive"] = []
        positivity["neutral"] = []


    for t in api.user_timeline(username, count=nb_tweet, include_
rts=False):
        if not detailled:
            positivity[get_sentiment(t.text)] += 1
        else:
            positivity[get_sentiment(t.text)].append(t.text)

    return positivity
```

In [64]:
```
print(get_profile_positivity("realDonaldTrump", 5000, detailled=True))
```

{'negative': ['Sleepy Eyes Chuck Todd is so happy with the fake v
oter tabulation process that he can’t even get the words out stra
ight. Sad to watch!', 'The Vice President has the power to reject
fraudulently chosen electors.', 'How can you certify an election
when the numbers being certified are verifiably WRONG. You will s
ee the real number… https://t.co/jfBOEEVjX7', 'Sorry, but the num
ber of votes in the Swing States that we are talking about is VER
Y LARGE and totally OUTCOME DETE… https://t.co/KZKiATT1lB', 'Why
haven’t they done signature verification in Fulton County, Georgi
a. Why haven’t they deducted all of the dead p… https://t.co/Bkz8
kFz41u', 'Some States are very slow to inoculate recipients despi
te successful and very large scale distribution of vaccines… http
s://t.co/uCpPoqzWuA', 'Our Republican Senate just missed the oppo
rtunity to get rid of Section 230, which gives unlimited power to
Big Tec… https://t.co/bevkn4zsNf', 'Watching @FoxNews is almost a
s bad as watching Fake News @CNN. New alternatives are developin
g!', 'New Lott study estimates 11,350 absentee votes lost to Trum
p in Georgia. Another 289,000 "excess (fraudulent) votes… http
s://t.co/V6cH6HYifI', 'The Wall Street Journal’s very boring &am
p; incoherent Editorial fails to mention my big & easy wins i
n Texas, Florida,… https://t.co/JZaL1rH3FB', '....Can you imagine
if the Republicans stole a Presidential Election from the Democra
ts - All hell would break out.… https://t.co/v4WDkxludM', 'Weak a
nd tired Republican “leadership” will allow the bad Defense Bill
to pass. Say goodbye to VITAL Section 230 te… https://t.co/fzmmkQ
Wmsv'], 'neutral': ['These are the things and events that happen
when a sacred landslide election victory is so unceremoniously &a
mp; viciou… https://t.co/uYLdmiz5Ka', 'https://t.co/Pm2PKV0Fp3',
'I am asking for everyone at the U.S. Capitol to remain peaceful.
No violence! Remember, WE are the Party of Law & O… https://
t.co/XWZnbZWwze', 'Mike Pence didn’t have the courage to do what
should have been done to protect our Country and our Constitutio
n, gi… https://t.co/pNct6O9uuy', 'https://t.co/izItBeFE6G', 'Even
Mexico uses Voter I.D.', 'The States want to redo their votes. Th
ey found out they voted on a FRAUD. Legislatures never approved.
Let them do it. BE STRONG!', 'They just happened to find 50,000 b
allots late last night. The USA is embarrassed by fools. Our Elec
tion Process is… https://t.co/cLaYvIvzab', 'THE REPUBLICAN PARTY
AND, MORE IMPORTANTLY, OUR COUNTRY, NEEDS THE PRESIDENCY MORE THA
N EVER BEFORE - THE POWER OF THE VETO. STAY STRONG!', 'States wan
t to correct their votes, which they now know were based on irreg
ularities and fraud, plus corrupt proces… https://t.co/i1ZARUmtHr
', 'If Vice President @Mike_Pence comes through for us, we will w
in the Presidency. Many States want to decertify the m… https://
t.co/9snX9u3SrA', 'Get smart Republicans. FIGHT! https://t.co/3fs
1oPVnAx', 'Just happened to have found another 4000 ballots from
Fulton County. Here we go!', 'Looks like they are setting up a bi
g “voter dump” against the Republican candidates. Waiting to see
how many votes they need?', 'BIG NEWS IN PENNSYLVANIA! https://t.
co/7JqTWYUgOr', 'I will be speaking at the SAVE AMERICA RALLY tom
orrow on the Ellipse at 11AM Eastern. Arrive early – doors open a
t… https://t.co/0OjqxuqF3Q', 'Antifa is a Terrorist Organization,
stay out of Washington. Law enforcement is watching you very clos
ely!… https://t.co/v0zJSzt2BI', 'Washington is being inundated wi
th people who don’t want to see an election victory stolen by emb
oldened Radical Le… https://t.co/u4LEyfrt5W', 'GEORGIA! Get out t
oday and VOTE for @KLoeffler and @Perduesenate! https://t.co/YKiS
x7d7lp', 'Reports are coming out of the 12th Congressional Distri
ct of Georgia that Dominion Machines are not working in cert… htt

ps://t.co/TdklHEmTnx', 'See you in D.C. https://t.co/ti4bChnPKz', 'https://t.co/tJkmEhSqY5', 'https://t.co/X0CIj5dFlV', 'https://t.co/zkFfhX0hOQ', 'https://t.co/gWj2obEKm8', 'https://t.co/9fRvjIgqKE', 'On my way, see you soon! https://t.co/7QW23k5b9r', 'https://t.co/3LNZWTO8NY', 'Heading to Georgia now. See you soon!', 'The “Surrender Caucus” within the Republican Party will go down in infamy as weak and ineffective “guardians” of ou… https://t.co/gDEuykZxdB', '“We are not acting to thwart the Democratic process, we are acting to protect it.” @SenRonJohnson', 'https://t.co/PXNO26lpoZ', 'https://t.co/2rLJfoKmSS', 'https://t.co/tuGfzK3QPE', 'https://t.co/JEzkcjgLSy', 'https://t.co/NmdLNsfDU2', 'https://t.co/dN2P40kJLW', '....@SenTedCruz @HawleyMO @Jim_Jordan @senatemajldr @GOPLeader & THE WORLD!', 'The Swing States did not even come close to following the dictates of their State Legislatures. These States “elec… https://t.co/6NsOU46HF6', 'Trump Speaks to State Legislators on Call About Decertifying Election https://t.co/z6BgCAe3zX via @BreitbartNews', 'https://t.co/FDUVk8cm9S', 'I will be there. Historic day! https://t.co/k6LStsWpfy', '“Georgia election data, just revealed, shows that over 17,000 votes illegally flipped from Trump to Biden.” @OANN… https://t.co/Meq7ioowNB', 'I spoke to Secretary of State Brad Raffensperger yesterday about Fulton County and voter fraud in Georgia. He was u… https://t.co/crqOHxwPtq', 'Republicans in Georgia must be careful of the political corruption in Fulton County, which is rampant. The Governor… https://t.co/ZKjiUIFD7l', 'The number of cases and deaths of the China Virus is far exaggerated in the United States because of @CDCgov’s ridi… https://t.co/6B8A4OOW1e', 'The vaccines are being delivered to the states by the Federal Government far faster than they can be administered!', 'GOP Senators Join Hawley in Objecting to Electoral College Votes https://t.co/f0JHfN4UUb via @BreitbartNews', 'https://t.co/moYUIcRq56', 'https://t.co/JLq7hHI42o', 'https://t.co/nslWcFwkCj', 'An attempt to steal a landslide win. Can’t let it happen! https://t.co/sKn4iTjUy0', '.@senatemajldr Mitch M, and all! https://t.co/zLIEKzfpFv', 'Civil War: Tucker Carlson Hits His Own Network in Epic Post-Election Monologue https://t.co/xUSQQWCa8q', '....Just a small portion of these votes give US a big and conclusive win in Georgia. Have they illegally destroyed… https://t.co/13AvDqYADp', 'TRANSPARENCY in medical pricing will be one of the biggest and most important things done for the American citizen.… https://t.co/YZGY38AYeV', 'For historical purposes remember, I was able to get rid of the INDIVIDUAL MANDATE, the most unpopular and expensive… https://t.co/JcQ7TodsDV', 'https://t.co/t6nWHJjJAN', '....Please remember who got it done!!!', 'Because of the Trump Administration, hospitals are now required, effective immediately, to publish their REAL PRICE… https://t.co/mqYtZXlBjP', 'Only because Biden got very few votes, just like the Election! https://t.co/sNIAJ1i5hu', 'Herschel is speaking the truth! https://t.co/6x9VLsc9qf', 'Republicans should have gotten rid of Section 230 in the Defense Bill, and you wouldn’t have had this problem. Neve… https://t.co/KZmam0zhAZ', 'January 6th. See you in D.C. https://t.co/vynZTv9lHb', '....is therefore both illegal and invalid, and that would include the two current Senatorial Elections. In Wisconsi… https://t.co/26IjrqlUOp', '....changes made to the voting process, rules and regulations, many made hastily before the election, and therefore… https://t.co/9OnMCZAn4e', 'Before even discussing the massive corruption which took place in the 2020 Election, which gives us far more votes… https://t.co/2seIXXRoIl', '.@FoxNews Weekend Daytime is not watchable. Switching over to @OANN!', 'NOW! https://t.co/hjC1WW64O6', 'Massive amounts of evidence will be presented on the 6th. We won, BIG! https://t.co/y

mncRrNR5t', 'The BIG Protest Rally in Washington, D.C., will take place at 11.00 A.M. on January 6th. Locational details to follow. StopTheSteal!', 'https://t.co/RBgT5SMWGT', 'https://t.co/7YHKqgjAub', 'https://t.co/FDUVk8cm9S', 'Sen. Josh Hawley Slams Walmart Tweet Calling Him a ‘Sore Loser‘ https://t.co/wmShnPcrzj via @BreitbartNews America… https://t.co/H38fgKR5BS', 'Finished off the year with the highest Stock Market in history. Setting records with your 401k’s, just like I said… https://t.co/v4mhFx0TdM', 'https://t.co/2kYtZF8Mei', 'https://t.co/S6fKpzoXZZ', 'We now have far more votes than needed to flip Georgia in the Presidential race. Massive VOTER FRAUD took place. Th… https://t.co/9xjslmlGf4', '.@BrianKempGA, his puppet Lt. Governor @GeoffDuncanGA, and Secretary of State, are disasters for Georgia. Won’t let… https://t.co/AzUchjw5uN', 'The United States had more votes than it had people voting, by a lot. This travesty cannot be allowed to stand. It… https://t.co/dH1j93GtiD', 'JANUARY SIXTH, SEE YOU IN DC!', 'Twitter is shadow banning like never before. A disgrace that our weak and ineffective political leadership refuses… https://t.co/bZPpwdsODN', 'Hearings from Atlanta on the Georgia Election overturn now being broadcast LIVE via @RSBNetwork! \nhttps://t.co/ogBvLbKfqG', 'Hearings from Atlanta on the Georgia Election overturn now being broadcast. Check it out. @OANN @newsmax and many m… https://t.co/Ejjc4SEieM', '$2000 ASAP!', 'The Federal Government has distributed the vaccines to the states. Now it is up to the states to administer. Get moving!', '...millions more votes than Trump, but can’t get anywhere close to him in this poll. No incoming president has ever… https://t.co/XDTBmSlzbH', '“Barack Obama was toppled from the top spot and President Trump claimed the title of the year’s Most Admired Man. T… https://t.co/aKtP3lCSpz', '....other acts of fraud and irregularities as well. STAY TUNED!', '...The consent decree signed by the “Secretary”, with the consent of Kemp, is perhaps even more poorly negotiated t… https://t.co/gfDa0RcV2V', 'https://t.co/Dcm41ThF5k', 'When are we going to be allowed to do signature verification in Fulton County, Georgia? The process is going VERY s… https://t.co/gkDts3Q0n7', 'It is up to the States to distribute the vaccines once brought to the designated areas by the Federal Government. W… https://t.co/eg0GJTW71l', 'Loeffler, Perdue Support Increasing Relief Payments to $2K https://t.co/xLFHNDOBIR via @BreitbartNews. Republicans… https://t.co/WukNDe4R4z', 'Unless Republicans have a death wish, and it is also the right thing to do, they must approve the $2000 payments AS… https://t.co/Ur6BCexxt1', '....and Congressmen/Congresswomen Elected. I do believe they forgot!', '...more votes than is needed by me to win Pennsylvania, not to mention hundreds of thousands of votes in other cate… https://t.co/Uo18Qad6JE', '“A group of Republican lawmakers in Pennsylvania say 200,000 more votes were counted in the 2020 Election than vote… https://t.co/onHvrHbUpN', '....being removed and brought home from foreign lands who do NOTHING for us. A disgraceful act of cowardice and tot… https://t.co/CsscO7MIj6', 'Give the people $2000, not $600. They have suffered enough! https://t.co/2jOVCnGtXS', 'This Tweet from @realDonaldTrump has been withheld in response to a report from the copyright holder. Learn more.', '“Breaking News: In Pennsylvania there were 205,000 more votes than there were voters. This alone flips the state to President Trump.”'], 'positive': ['Please support our Capitol Police and Law Enforcement. They are truly on the side of our Country. Stay peaceful!', 'These scoundrels are only toying with the @sendavidperdue (a great guy) vote. Just didn’t want to announce quite ye… https://t.co/PWkWMR8npN', 'I hope the Democrats, and even more importantly, the weak and ineffective RINO section of the Re

```
publican Party, are… https://t.co/qOXn3ASfvH', 'Georgia, get out
and VOTE for two great Senators, @KLoeffler and @sendavidperdue.
So important to do so!', 'Pleased to announce that @KLoeffler &am
p; @sendavidperdue have just joined our great #StopTheSteal group
of Senators. T… https://t.co/rBcAsw9bx6', '“We’ve seen in the las
t few months, unprecedented amounts of Voter Fraud.” @SenTedCruz
True!', 'Great! https://t.co/f0nKNPgqWg', 'Something how Dr. Fauc
i is revered by the LameStream Media as such a great professiona
l, having done, they say, suc… https://t.co/xEpEJBWmKu', 'So tru
e. Thanks Josh! https://t.co/lacUQC6IHh', '...And after they see
the facts, plenty more to come...Our Country will love them for i
t! #StopTheSteal https://t.co/0IdbiACLIb', 'Wow, I guess it’s not
good to go against a President who everyone in Georgia knows got
you into office! https://t.co/4xUNdOncoB', 'MAKE AMERICA GREAT AG
AIN!', 'Will be in Georgia on Monday night, 9:00 P.M. to RALLY fo
r two GREAT people, @sendavidperdue &amp; @KLoeffler. GET READY T
O VOTE ON TUESDAY!!!', 'Thank you Madison! https://t.co/XyaAHO9Cw
C', 'A great honor! https://t.co/U4WFBWrnF7', 'I hope to see the
great Governor of South Dakota @KristiNoem, run against RINO @Sen
JohnThune, in the upcoming 2022… https://t.co/3z3zTYTvH7', 'HAPPY
NEW YEAR!', 'Thank you, a great honor! https://t.co/9GfT5c8hY0',
'....that, quite frankly, didn’t have much of a chance, like 7, 8
or 9. The Presidential Election was Rigged with hu… https://t.co/
56MgHJkCYU', 'I love the Great State of Georgia, but the people w
ho run it, from the Governor, @BrianKempGA, to the Secretary of…
https://t.co/HL2Fh8oO7F', '$2000 for our great people, not $600!
```

In [60]:
```
print(get_profile_positivity("the_weird_weeb", 5000))
```

```
{'negative': 5, 'neutral': 65, 'positive': 1}
```