

# Twitter\_Opinion\_AQV

June 6, 2017

## 1 ANALYSE DES OPINIONS SOUS TWITTER

Audrey Quessada  
INF344

```
In [1]: %matplotlib inline
import os
import csv as csv
import nltk as nltk
import re
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from collections import Counter
from IPython.display import display, HTML
from nltk.corpus import wordnet as wn
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sentiwordnet import SentiWordNetCorpusReader, SentiSynset

In [2]: csvTweetData = 'testdata.manual.2009.06.14.csv'

In [3]: colonne = ['Notes_tweet', 'Index', 'Date', 'Subject', 'Author', 'Tweet']
df1 = pd.read_csv(csvTweetData, sep=',', header=None)
df1.columns = colonne
N, _ = df1.shape
print("Il y a %d tweets" %N)
print(display(df1.iloc[0:10]))
```

Il y a 498 tweets

	Notes_tweet	Index	Date	Subject	Author	\
0	4	3	Mon May 11 03:17:40 UTC 2009	kindle2	tpryan	
1	4	4	Mon May 11 03:18:03 UTC 2009	kindle2	vcu451	
2	4	5	Mon May 11 03:18:54 UTC 2009	kindle2	chadfu	

3	4	6	Mon May 11 03:19:04 UTC 2009	kindle2	SIX15
4	4	7	Mon May 11 03:21:41 UTC 2009	kindle2	yamarama
5	4	8	Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme
6	0	9	Mon May 11 03:22:30 UTC 2009	aig	Seth937
7	4	10	Mon May 11 03:26:10 UTC 2009	jquery	dcostalis
8	4	11	Mon May 11 03:27:15 UTC 2009	twitter	PJ_King
9	4	12	Mon May 11 03:29:20 UTC 2009	obama	mandanicole

```

                                Tweet
0 @stellargirl I loooooooooovvvvvveee my Kindle2. ...
1 Reading my kindle2... Love it... Lee childs i...
2 Ok, first assesment of the #kindle2 ...it fuck...
3 @kenburbary You'll love your Kindle2. I've had...
4 @mikefish Fair enough. But i have the Kindle2...
5 @richardebaker no. it is too big. I'm quite ha...
6 Fuck this economy. I hate aig and their non lo...
7                                JQuery is my new best friend.
8                                Loves twitter
9 how can you not love Obama? he makes jokes abo...

```

None

## 1.1 1.1Pre-processing simple des tweets

Import du dictionnaire abréviation (<http://perso.telecom-paristech.fr/~clavel/DonneesTweets/SlangLookupTable.txt>)

Le fichier text est converti en dictionnaire python où pour chaque abréviation en clé correspond sa traduction en valeur.

```

In [4]: f = open('./dico_slang.txt')
        slanglines = f.readlines()
        f.close()

In [5]: list_slang = [x.replace('\t', ' ').replace('\n', '') for x in slanglines]
        #print(list_slang[0].split())
        #print(list_slang)
        slang = [l.split()[0] for l in list_slang if l]
        translate_slang = [' '.join(l.split()[1:]) for l in list_slang if l]
        dico_slang = {}
        for i, j in zip(slang, translate_slang):
            dico_slang[i] = j
        print(dico_slang)

{'wfm': 'works for me', 'ru': 'are you', 'ttyl': 'talk to you later', 'thx': 'thank

```

Extraction des destinataires, des urls, des hashtags et nettoyages des tweets:  
On crée un dictionnaire qui va contenir les adresses url, les destinataires des tweets et les hashtags

répertoriés.

Pour cela on va utiliser plusieurs regex nous permettant de récupérer ces informations.

On va aussi “nettoyer” les tweets, à savoir, une fois qu’on a enlever les champs sus-mentionnés, on met tous les tweets en minuscule et on enlève la ponctuation ou certains caractères spéciaux.

Il faut enlever RT qui veut dire retweet (ou on peut compter le nombre de retweet)

On va tester dans un premier temps les regex

```
In [6]: test0 = "Blah, blah, blah same old same old. No plans today, going back to
test1 = 'Nike owns NBA Playoffs ads w/ LeBron, Kobe, Carmelo? http://ow.ly/
test2 = '@phyreman9 Google is always a good place to look. Should\'ve menti
test3 = 'Life?s a bitch? and so is Dick Cheney. #p2 #bipart #tlot #tcot #hh
test4 = '@anna_debenham what was the php jquery hack?'
test5 = '@ work til 6pm... lets go lakers!!!'
test6 = "RT @PassionModel 'American Idol' fashion: Adam Lambert tones down,
test7 = "RT @justindavey: RT @tweetmeme GM OnStar now instantly sends accid
test8 = 'RT @shrop: Awesome JQuery reference book for Coda! http://www.macp
test9 = "Time Warner Cable Pulls the Plug on 'The Girlfriend Experience' -
test10 = 'House Correspondents dinner was last night whoopi, barbara &
test11 = "I'm listening to 'P.Y.T' by Danny Gokey &3 &3 &3 Aww, he
```

On va écrire des fonctions qui vont permettre de nettoyer les tweets

```
In [7]: def get_url(string):
        '''fonction qui permet de trouver tous les liens, qui prend en entrée u
        retourne en sortie la liste des url et la chaine de caractères nettoyée
        reg1 = re.findall("http\S+",string)
        reg2 = re.findall('([\da-z\.-]+) (\.[a-z\.]{2,6}\\/\w+)', string)
        if reg1:
            return ''.join(reg1[0]), re.sub("http\S+", '',string)
        elif reg2:
            return ''.join(reg2[0]), re.sub('([\da-z\.-]+) (\.[a-z\.]{2,6}\\/\w+)'
        else:
            return [], string
```

```
In [8]: print(get_url(test6))
        print(' ')
        print(get_url(test9))
        print(' ')
        print(get_url(test0))
```

```
('http://cli.gs/7JWSHV', "RT @PassionModel 'American Idol' fashion: Adam Lambert to
```

```
('www.tinyurl.com/m595fk', "Time Warner Cable Pulls the Plug on 'The Girlfriend Exp
```

```
([], 'Blah, blah, blah same old same old. No plans today, going back to sleep I gue
```

```
In [9]: def get_hashtag(string):
        reg = re.findall('#(\w+)', string)
```

```

        if reg:
            return reg, re.sub('#(\w+)', '', string, flags = re.UNICODE)
        else:
            return [], string

In [10]: print(get_hashtag(test1))
          print(' ')
          print(get_hashtag(test0))

(['Adidas', 'Billups', 'Howard', 'Marketing', 'Branding'], 'Nike owns NBA Playoffs')

([], 'Blah, blah, blah same old same old. No plans today, going back to sleep I guess')

In [11]: def get_at(string):
          reg = re.findall('@(\w+)', string)
          if reg:
              return reg, re.sub('@', '', string, flags = re.UNICODE)
          else:
              return [], string

In [12]: print(get_at(test2))
          print(' ')
          print(get_at(test0))
          print(' ')
          print(get_at(test5))

(['phyreman9', 'KimbleT'], "phyreman9 Google is always a good place to look. Shouldn't you?")

([], 'Blah, blah, blah same old same old. No plans today, going back to sleep I guess')

([], '@ work til 6pm... lets go lakers!!!')

In [13]: def get_retweet(string):
          reg = re.findall('RT', string)
          if reg:
              return reg, re.sub('RT', '', string)
          else:
              return [], string

In [14]: print(get_retweet(test7))
          print(' ')
          print(get_retweet(test0))

(['RT', 'RT'], ' @justindavey: @tweetmeme GM OnStar now instantly sends accident location')

([], 'Blah, blah, blah same old same old. No plans today, going back to sleep I guess')

```



```

        list_RT.extend(rt)
    else:
        pass

    if destinataire == True:
        at, tweet = get_at(tweet)
        list_destinataire.extend(at)
        at_per_tweet.append(at)
    else:
        pass

    if hashtag == True:
        hasht, tweet = get_hashtag(tweet)
        hash_per_tweet.append(hasht)
        list_hashtag.extend(hasht)
    else:
        pass

    if url == True:
        ur, tweet = get_url(tweet)
        list_url.extend(ur)
        url_per_tweet.append(ur)
    else:
        pass

    if punct == True:
        tweet = get_tweet_cleant(tweet)
    else:
        pass

    clean_tweet.append(tweet)

    dico_tweet['Destinataire'] = at_per_tweet
    dico_tweet['hashtag'] = hash_per_tweet
    dico_tweet['url'] = url_per_tweet

    n_hashtag = len(list_hashtag)
    n_destinataire = len(list_destinataire)
    n_retweet = len(list_RT)
    n_url = len(list_url)

    return n_url, n_hashtag, n_destinataire, n_retweet, dico_tweet, clean_tweet

```

```

In [18]: n_url, n_hashtag, n_destinataire, n_retweet, dico_tweet, clean_tweet = app
print('Nombre d\'url: ', n_url)
print('Nombre de hashtags: ', n_hashtag)
print('Nombre de destinataires: ', n_destinataire)
print('Nombre de retweet: ', n_retweet)

```

Уш10 'got' Ghome'sednechickhi fihto bhebostrwvseset Dayukimbeht', Noshahqonday Dxtwipidobakepwarmee a2da

Уш10 'got' Ghome'sednechickhi fihto bhebostrwvseset Dayukimbeht', Noshahqonday Dxtwipidobakepwarmee a2da

Уш10 'got' Ghome'sednechickhi fihto bhebostrwvseset Dayukimbeht', Noshahqonday Dxtwipidobakepwarmee a2da

Уш10 'got' Ghome'sednechickhi fihto bhebostrwvseset Dayukimbeht', Noshahqonday Dxtwipidobakepwarmee a2da

```

33             []             []             []
34             []             []             []
35             []             []             []
36             []             []             []
37     [sekseemess]           []             []
38     [jessverr]             []     http://bit.ly/13t7NR
39     [spinuzzi]             []             []

```

None

```

In [20]: def get_token_from_tweet(list_tweet, lower=True):
         tkn = nltk.word_tokenize
         if lower == True:
             return [tkn(x.lower()) for x in list_tweet]
         else:
             return [tkn(x) for x in list_tweet]

```

```

In [21]: words = get_token_from_tweet(clean_tweet, lower=False)
         print(words)

```

```

[['stellargirl', 'I', 'loooooooooovvvvvveee', 'my', 'Kindle2', 'Not', 'that', 'the',

```



	Notes_tweet	Index		Date	Subject	Author	\
0	4	3	Mon May 11 03:17:40 UTC 2009	kindle2	tpryan		
1	4	4	Mon May 11 03:18:03 UTC 2009	kindle2	vcu451		
2	4	5	Mon May 11 03:18:54 UTC 2009	kindle2	chadfu		
3	4	6	Mon May 11 03:19:04 UTC 2009	kindle2	SIX15		
4	4	7	Mon May 11 03:21:41 UTC 2009	kindle2	yamarama		
5	4	8	Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme		
6	0	9	Mon May 11 03:22:30 UTC 2009	aig	Seth937		
7	4	10	Mon May 11 03:26:10 UTC 2009	jquery	dcostalis		
8	4	11	Mon May 11 03:27:15 UTC 2009	twitter	PJ_King		
9	4	12	Mon May 11 03:29:20 UTC 2009	obama	mandanicole		
10	2	13	Mon May 11 03:32:42 UTC 2009	obama	jpeb		
11	0	14	Mon May 11 03:32:48 UTC 2009	obama	kylesellers		
12	4	15	Mon May 11 03:33:38 UTC 2009	obama	theviewfans		
13	4	16	Mon May 11 05:05:58 UTC 2009	nike	MumsFP		
14	0	17	Mon May 11 05:06:22 UTC 2009	nike	vincentx24x		
15	4	18	Mon May 11 05:20:15 UTC 2009	lebron	cameronwylie		
16	0	19	Mon May 11 05:20:28 UTC 2009	lebron	luv8242		
17	4	20	Mon May 11 05:21:04 UTC 2009	lebron	mtgillikin		
18	0	21	Mon May 11 05:21:37 UTC 2009	lebron	ursecretdezire		
19	4	22	Mon May 11 05:21:45 UTC 2009	lebron	Native_01		

	Tweet	\
0	@stellargirl I loooooooooovvvvvveee my Kindle2. ...	
1	Reading my kindle2... Love it... Lee childs i...	
2	Ok, first assesment of the #kindle2 ...it fuck...	
3	@kenburbary You'll love your Kindle2. I've had...	
4	@mikefish Fair enough. But i have the Kindle2...	
5	@richardebaker no. it is too big. I'm quite ha...	
6	Fuck this economy. I hate aig and their non lo...	
7	jQuery is my new best friend.	
8	Loves twitter	
9	how can you not love Obama? he makes jokes abo...	
10	Check this video out -- President Obama at the...	
11	@Karoli I firmly believe that Obama/Pelosi hav...	
12	House Correspondents dinner was last night who...	
13	Watchin Espn..Jus seen this new Nike Commerica...	
14	dear nike, stop with the flywire. that shit is...	
15	#lebron best athlete of our generation, if not...	
16	I was talking to this guy last night and he wa...	
17	i love lebron. http://bit.ly/PdHur	
18	@ludaj Juice LeBron is a Beast, but I'm still ch...	
19	@Pmillzz lebron IS THE BOSS	

	clean_tweet	\
0	stellargirl I love my Kindle2 Not that the DX ...	
1	Reading my kindle2 Love it Lee childs is good ...	
2	Ok first assesment of the it fucking rocks	

```

3 kenburbary You ll love your Kindle2 I ve had m...
4 mikefish Fair enough But i have the Kindle2 an...
5 richardebaker no it is too big I m quite happy...
6 Fuck this economy I hate aig and their non loa...
7           JQuery is my new best friend
8           Loves twitter
9 how can you not love Obama he makes jokes abou...
10 Check this video out -- President Obama at the...
11 Karoli I firmly believe that Obama/Pelosi have...
12 House Correspondents dinner was last night who...
13 Watchin Espn Jus seen this new Nike Commerical...
14 dear nike stop with the flywire that shit is a...
15 best athlete of our generation if not all time...
16 I was talking to this guy last night and he wa...
17           i love lebron
18 ludajoice Lebron is a Beast but I m still chee...
19           Pmillzz lebron IS THE BOSS

```

```

tokens
0 [stellargirl, I, love, my, Kindle2, Not, that,...
1 [Reading, my, kindle2, Love, it, Lee, childs, ...
2 [Ok, first, assesment, of, the, it, fucking, r...
3 [kenburbary, You, ll, love, your, Kindle2, I, ...
4 [mikefish, Fair, enough, But, i, have, the, Ki...
5 [richardebaker, no, it, is, too, big, I, m, qu...
6 [Fuck, this, economy, I, hate, aig, and, their...
7           [Jquery, is, my, new, best, friend]
8           [Loves, twitter]
9 [how, can, you, not, love, Obama, he, makes, j...
10 [Check, this, video, out, --, President, Obama...
11 [Karoli, I, firmly, believe, that, Obama/Pelos...
12 [House, Correspondents, dinner, was, last, nig...
13 [Watchin, Espn, Jus, seen, this, new, Nike, Co...
14 [dear, nike, stop, with, the, flywire, that, s...
15 [best, athlete, of, our, generation, if, not, ...
16 [I, was, talking, to, this, guy, last, night, ...
17           [i, love, lebron]
18 [ludajoice, Lebron, is, a, Beast, but, I, m, s...
19           [Pmillzz, lebron, IS, THE, BOSS]

```

None

Etiquetage grammatical:  
on construit un dictionnaire pour chaque tweet contenant tous les POS-tags

```

In [25]: tag = nltk.pos_tag
         def getPOS(list_token):

```

```

all_tag = []
for i in range(len(list_token)):
    all_tag.append(tag(list_token[i]))
return all_tag

```

```

In [26]: tag_list_tot = getPOS(words_noslang)
df1['POS'] = tag_list_tot
print(display(df1.iloc[0:20]))

```

	Notes_tweet	Index		Date	Subject	Author \
0	4	3	Mon May 11 03:17:40 UTC 2009	kindle2	tpryan	
1	4	4	Mon May 11 03:18:03 UTC 2009	kindle2	vcu451	
2	4	5	Mon May 11 03:18:54 UTC 2009	kindle2	chadfu	
3	4	6	Mon May 11 03:19:04 UTC 2009	kindle2	SIX15	
4	4	7	Mon May 11 03:21:41 UTC 2009	kindle2	yamarama	
5	4	8	Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme	
6	0	9	Mon May 11 03:22:30 UTC 2009	aig	Seth937	
7	4	10	Mon May 11 03:26:10 UTC 2009	jquery	dcostalis	
8	4	11	Mon May 11 03:27:15 UTC 2009	twitter	PJ_King	
9	4	12	Mon May 11 03:29:20 UTC 2009	obama	mandanicole	
10	2	13	Mon May 11 03:32:42 UTC 2009	obama	jpeb	
11	0	14	Mon May 11 03:32:48 UTC 2009	obama	kylesellers	
12	4	15	Mon May 11 03:33:38 UTC 2009	obama	theviewfans	
13	4	16	Mon May 11 05:05:58 UTC 2009	nike	MumsFP	
14	0	17	Mon May 11 05:06:22 UTC 2009	nike	vincentx24x	
15	4	18	Mon May 11 05:20:15 UTC 2009	lebron	cameronwylie	
16	0	19	Mon May 11 05:20:28 UTC 2009	lebron	luv8242	
17	4	20	Mon May 11 05:21:04 UTC 2009	lebron	mtgillikin	
18	0	21	Mon May 11 05:21:37 UTC 2009	lebron	ursecretdezire	
19	4	22	Mon May 11 05:21:45 UTC 2009	lebron	Native_01	

	Tweet \
0	@stellargirl I loooooooooovvvvvveee my Kindle2. ...
1	Reading my kindle2... Love it... Lee childs i...
2	Ok, first assesment of the #kindle2 ...it fuck...
3	@kenburbary You'll love your Kindle2. I've had...
4	@mikefish Fair enough. But i have the Kindle2...
5	@richardebaker no. it is too big. I'm quite ha...
6	Fuck this economy. I hate aig and their non lo...
7	Jquery is my new best friend.
8	Loves twitter
9	how can you not love Obama? he makes jokes abo...
10	Check this video out -- President Obama at the...
11	@KarloI I firmly believe that Obama/Pelosi hav...
12	House Correspondents dinner was last night who...
13	Watchin Espn..Jus seen this new Nike Commerica...
14	dear nike, stop with the flywire. that shit is...
15	#lebron best athlete of our generation, if not...

16 I was talking to this guy last night and he wa...  
 17 i love lebron. http://bit.ly/PdHur  
 18 @ludaj Juice Lebron is a Beast, but I'm still ch...  
 19 @Pmillzz lebron IS THE BOSS

clean\_tweet \

0 stellargirl I love my Kindle2 Not that the DX ...  
 1 Reading my kindle2 Love it Lee childs is good ...  
 2 Ok first assesment of the it fucking rocks  
 3 kenburbary You ll love your Kindle2 I ve had m...  
 4 mikefish Fair enough But i have the Kindle2 an...  
 5 richardebaker no it is too big I m quite happy...  
 6 Fuck this economy I hate aig and their non loa...  
 7 JQuery is my new best friend  
 8 Loves twitter  
 9 how can you not love Obama he makes jokes abou...  
 10 Check this video out -- President Obama at the...  
 11 Karoli I firmly believe that Obama/Pelosi have...  
 12 House Correspondents dinner was last night who...  
 13 Watchin Espn Jus seen this new Nike Commerical...  
 14 dear nike stop with the flywire that shit is a...  
 15 best athlete of our generation if not all time...  
 16 I was talking to this guy last night and he wa...  
 17 i love lebron  
 18 ludaj Juice Lebron is a Beast but I m still chee...  
 19 Pmillzz lebron IS THE BOSS

tokens \

0 [stellargirl, I, love, my, Kindle2, Not, that,...  
 1 [Reading, my, kindle2, Love, it, Lee, childs, ...  
 2 [Ok, first, assesment, of, the, it, fucking, r...  
 3 [kenburbary, You, ll, love, your, Kindle2, I, ...  
 4 [mikefish, Fair, enough, But, i, have, the, Ki...  
 5 [richardebaker, no, it, is, too, big, I, m, qu...  
 6 [Fuck, this, economy, I, hate, aig, and, their...  
 7 [Jquery, is, my, new, best, friend]  
 8 [Loves, twitter]  
 9 [how, can, you, not, love, Obama, he, makes, j...  
 10 [Check, this, video, out, --, President, Obama...  
 11 [Karoli, I, firmly, believe, that, Obama/Pelos...  
 12 [House, Correspondents, dinner, was, last, nig...  
 13 [Watchin, Espn, Jus, seen, this, new, Nike, Co...  
 14 [dear, nike, stop, with, the, flywire, that, s...  
 15 [best, athlete, of, our, generation, if, not, ...  
 16 [I, was, talking, to, this, guy, last, night, ...  
 17 [i, love, lebron]  
 18 [ludaj Juice, Lebron, is, a, Beast, but, I, m, s...  
 19 [Pmillzz, lebron, IS, THE, BOSS]

```

                                POS
0  [(stellargirl, NN), (I, PRP), (love, VBP), (my...
1  [(Reading, VBG), (my, PRP$), (kindle2, NN), (L...
2  [(Ok, NNP), (first, JJ), (assesment, NN), (of,...
3  [(kenburbary, NN), (You, PRP), (ll, VBP), (lov...
4  [(mikefish, JJ), (Fair, NNP), (enough, RB), (B...
5  [(richardebaker, NN), (no, DT), (it, PRP), (is...
6  [(Fuck, NNP), (this, DT), (economy, NN), (I, P...
7  [(Jquery, NN), (is, VBZ), (my, PRP$), (new, JJ...
8  [(Loves, NNS), (twitter, NN)]
9  [(how, WRB), (can, MD), (you, PRP), (not, RB),...
10 [(Check, VB), (this, DT), (video, NN), (out, R...
11 [(Karloli, NNP), (I, PRP), (firmly, RB), (belie...
12 [(House, NNP), (Correspondents, NNP), (dinner,...
13 [(Watchin, NNP), (Espn, NNP), (Jus, NNP), (see...
14 [(dear, JJ), (nike, IN), (stop, NN), (with, IN...
15 [(best, JJS), (athlete, NN), (of, IN), (our, P...
16 [(I, PRP), (was, VBD), (talking, VBG), (to, TO...
17 [(i, NN), (love, VBP), (lebron, NN)]
18 [(ludaj Juice, NN), (Lebron, NNP), (is, VBZ), (a...
19 [(Pmillzz, NNP), (lebron, NN), (IS, VBZ), (THE...

```

None

Les verbes sont repérés par les tags VB, VBZ , VBP, VBN, VBD, MD

```

In [27]: def get_verbs(list_tag):
          tot_pos = []
          for i in range (len(list_tag)):
              for j in range (len(list_tag[i])):
                  tot_pos.append(list_tag[i][j][1])
          count_VB = Counter(tot_pos)
          verb_tag = ['VB', 'VBZ' , 'VBP', 'VBG', 'VBN', 'VBD', 'MD']
          sum_vb = 0
          for k,v in count_VB.items():
              if k in verb_tag:
                  sum_vb += v
          return sum_vb, count_VB

```

```

In [28]: sum_vb, count_VB = get_verbs(tag_list_tot)
          print(count_VB)
          print(' ')
          print('Le nombre total de POS verbes sur l\'ensemble des tweets est: ', su

```

```

Counter({'NNP': 1305, 'NN': 1118, 'IN': 579, 'DT': 435, 'JJ': 434, 'PRP': 373, 'RB'

```

Le nombre total de POS verbes sur l'ensemble des tweets est: 1170

## 1.2 1.2 Algorithme de détection v1 : appel au dictionnaire Sentiwordnet

```
In [29]: wn.synsets('dog')
```

```
Out[29]: [Synset('dog.n.01'),  
          Synset('frump.n.01'),  
          Synset('dog.n.03'),  
          Synset('cad.n.01'),  
          Synset('frank.n.02'),  
          Synset('pawl.n.01'),  
          Synset('andiron.n.01'),  
          Synset('chase.v.01')]
```

```
In [30]: swn_filename = './SentiWordNet_3.0.0_20130122.txt'  
swn = SentiWordNetCorpusReader(swn_filename)  
test = swn.senti_synset('breakdown.n.03')  
print('test d\'un mot: ', test.synset.name)  
print('score positif: ', test.pos_score)  
print('score negatif: ', test.neg_score)
```

```
test d'un mot: <bound method Synset.name of Synset('breakdown.n.03')>  
score positif: 0.0  
score negatif: 0.25
```

```
In [31]: noun_tag = ['NN', 'NNS', 'NNP', 'NNPS']  
adj_tag = ['JJ', 'JJR', 'JJS']  
adv_tag = ['RB', 'RBR', 'RBS']  
v_tag = ['VB', 'VBZ', 'VBP', 'VBG', 'VBN', 'VBD']  
all_tag = ['NN', 'NNS', 'NNP', 'NNPS', 'JJ', 'JJR', 'JJS', 'RB', 'RBR', 'RBS', 'VB', 'VBZ', 'VBP', 'VBG', 'VBN', 'VBD']
```

On récupère les POS-tags valides à savoir ceux qui correspondent aux noms, adjectifs, verbes, et adverbes.

```
In [32]: def wordnet_sanitize(word):  
        """  
        author:christopher potts  
        http://compprag.christopherpotts.net/wordnet.html  
        Ensure that word is a (string, pos) pair that WordNet can understand.  
  
        Argument: word (str, str) -- a (string, pos) pair  
  
        Value: a possibly modified (string, pos) pair, where pos=None if  
        the input pos is outside of WordNet.  
        """  
        string, tag = word
```

```

string = string.lower()
tag = tag.lower()
if tag.startswith('v'):    tag = 'v'
elif tag.startswith('n'):  tag = 'n'
elif tag.startswith('j'):  tag = 'a'
elif tag.startswith('rb'): tag = 'r'
if tag in ('a', 'n', 'r', 'v'):
    return (string, tag)
else:
    return (string, None)

```

```

In [33]: def get_score_senti(tag_list):
    tweet_sanitizate = []
    pos_score = []
    pos_sum_score = []
    neg_score = []
    neg_sum_score = []
    for ttag in tag_list:
        n = len(ttag)
        new_san = [wordnet_sanitizate(ttag[i]) for i in range(n) if ttag[i] != '']
        tweet_sanitizate.append(new_san)
    for tweet in tweet_sanitizate:
        pos_tweet = []
        neg_tweet = []
        for i in range(len(tweet)):
            w, t = tweet[i]
            tweet_senti = list(swn.senti_synsets(w,t))
            if tweet_senti:
                senti = list(swn.senti_synsets(w,t))[0]
                pos_tweet.append(senti.pos_score)
                neg_tweet.append(senti.neg_score)
        pos_score.append(pos_tweet)
        neg_score.append(neg_tweet)
        pos_sum_score.append(sum(pos_tweet))
        neg_sum_score.append(sum(neg_tweet))
    return tweet_sanitizate, pos_score, pos_sum_score, neg_score, neg_sum_score

```

```

In [34]: tweet_sanitizate, pos_score, pos_sum_score, neg_score, neg_sum_score = get_score_senti(tweet_list)
print(tweet_sanitizate[0])

```

```

[('stellargirl', 'n'), ('love', 'v'), ('kindle2', 'n'), ('not', 'r'), ('dx', 'n'), ('', '')]

```

```

In [35]: print('Score positif des 10 premiers tweets')
print(pos_sum_score[:10])
print('')
print('Score négatif des 10 premiers tweets')
print(neg_sum_score[:10])

```

Score positif des 10 premiers tweets

[1.25, 1.75, 0.375, 1.875, 1.0, 1.5, 0.0, 1.5, 0.625, 0.875]

Score négatif des 10 premiers tweets

[1.0, 0.125, 0.125, 1.875, 0.125, 1.125, 0.875, 0.125, 0.0, 0.625]

```
In [36]: df1['Score_Pos'] = pos_sum_score
df1['Score_Neg'] = neg_sum_score
note_global = []
for p,n in zip(pos_sum_score, neg_sum_score):
    if p > n:
        note_global.append(1)
    elif p == n:
        note_global.append(0)
    else:
        note_global.append(-1)
df1['Note_globale'] = note_global
print(display(df1.iloc[0:20]))
```

	Notes_tweet	Index		Date	Subject	Author \
0	4	3	Mon May 11 03:17:40 UTC 2009	kindle2	tpryan	
1	4	4	Mon May 11 03:18:03 UTC 2009	kindle2	vcu451	
2	4	5	Mon May 11 03:18:54 UTC 2009	kindle2	chadfu	
3	4	6	Mon May 11 03:19:04 UTC 2009	kindle2	SIX15	
4	4	7	Mon May 11 03:21:41 UTC 2009	kindle2	yamarama	
5	4	8	Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme	
6	0	9	Mon May 11 03:22:30 UTC 2009	aig	Seth937	
7	4	10	Mon May 11 03:26:10 UTC 2009	jquery	dcostalis	
8	4	11	Mon May 11 03:27:15 UTC 2009	twitter	PJ_King	
9	4	12	Mon May 11 03:29:20 UTC 2009	obama	mandanicole	
10	2	13	Mon May 11 03:32:42 UTC 2009	obama	jpeb	
11	0	14	Mon May 11 03:32:48 UTC 2009	obama	kylesellers	
12	4	15	Mon May 11 03:33:38 UTC 2009	obama	theviewfans	
13	4	16	Mon May 11 05:05:58 UTC 2009	nike	MumsFP	
14	0	17	Mon May 11 05:06:22 UTC 2009	nike	vincentx24x	
15	4	18	Mon May 11 05:20:15 UTC 2009	lebron	cameronwylie	
16	0	19	Mon May 11 05:20:28 UTC 2009	lebron	luv8242	
17	4	20	Mon May 11 05:21:04 UTC 2009	lebron	mtgillikin	
18	0	21	Mon May 11 05:21:37 UTC 2009	lebron	ursecretdezire	
19	4	22	Mon May 11 05:21:45 UTC 2009	lebron	Native_01	

	Tweet \
0	@stellargirl I loooooooooovvvvvveee my Kindle2. ...
1	Reading my kindle2... Love it... Lee childs i...
2	Ok, first assesment of the #kindle2 ...it fuck...
3	@kenburbary You'll love your Kindle2. I've had...
4	@mikefish Fair enough. But i have the Kindle2...



5 @richardebaker no. it is too big. I'm quite ha...  
 6 Fuck this economy. I hate aig and their non lo...  
 7 JQuery is my new best friend.  
 8 Loves twitter  
 9 how can you not love Obama? he makes jokes abo...  
 10 Check this video out -- President Obama at the...  
 11 @KarloI I firmly believe that Obama/Pelosi hav...  
 12 House Correspondents dinner was last night who...  
 13 Watchin Espn..Jus seen this new Nike Commerica...  
 14 dear nike, stop with the flywire. that shit is...  
 15 #lebron best athlete of our generation, if not...  
 16 I was talking to this guy last night and he wa...  
 17 i love lebron. <http://bit.ly/PdHur>  
 18 @ludajoice Lebron is a Beast, but I'm still ch...  
 19 @Pmillzz lebron IS THE BOSS

clean\_tweet \

0 stellargirl I love my Kindle2 Not that the DX ...  
 1 Reading my kindle2 Love it Lee childs is good ...  
 2 Ok first assesment of the it fucking rocks  
 3 kenburbary You ll love your Kindle2 I ve had m...  
 4 mikefish Fair enough But i have the Kindle2 an...  
 5 richardebaker no it is too big I m quite happy...  
 6 Fuck this economy I hate aig and their non loa...  
 7 JQuery is my new best friend  
 8 Loves twitter  
 9 how can you not love Obama he makes jokes abou...  
 10 Check this video out -- President Obama at the...  
 11 Karoli I firmly believe that Obama/Pelosi have...  
 12 House Correspondents dinner was last night who...  
 13 Watchin Espn Jus seen this new Nike Commerical...  
 14 dear nike stop with the flywire that shit is a...  
 15 best athlete of our generation if not all time...  
 16 I was talking to this guy last night and he wa...  
 17 i love lebron  
 18 ludajoice Lebron is a Beast but I m still chee...  
 19 Pmillzz lebron IS THE BOSS

tokens \

0 [stellargirl, I, love, my, Kindle2, Not, that,...  
 1 [Reading, my, kindle2, Love, it, Lee, childs, ...  
 2 [Ok, first, assesment, of, the, it, fucking, r...  
 3 [kenburbary, You, ll, love, your, Kindle2, I, ...  
 4 [mikefish, Fair, enough, But, i, have, the, Ki...  
 5 [richardebaker, no, it, is, too, big, I, m, qu...  
 6 [Fuck, this, economy, I, hate, aig, and, their...  
 7 [Jquery, is, my, new, best, friend]  
 8 [Loves, twitter]

9 [how, can, you, not, love, Obama, he, makes, j...  
 10 [Check, this, video, out, --, President, Obama...  
 11 [Karloli, I, firmly, believe, that, Obama/Pelos...  
 12 [House, Correspondents, dinner, was, last, nig...  
 13 [Watchin, Espn, Jus, seen, this, new, Nike, Co...  
 14 [dear, nike, stop, with, the, flywire, that, s...  
 15 [best, athlete, of, our, generation, if, not, ...  
 16 [I, was, talking, to, this, guy, last, night, ...  
 17 [i, love, lebron]  
 18 [ludajoice, Lebron, is, a, Beast, but, I, m, s...  
 19 [Pmillzz, lebron, IS, THE, BOSS]

	POS	Score_Pos	Score_Neg	\
0 [(stellargirl, NN), (I, PRP), (love, VBP), (my...		1.250	1.000	
1 [(Reading, VBG), (my, PRP\$), (kindle2, NN), (L...		1.750	0.125	
2 [(Ok, NNP), (first, JJ), (assesment, NN), (of,...		0.375	0.125	
3 [(kenburbary, NN), (You, PRP), (ll, VBP), (lov...		1.875	1.875	
4 [(mikefish, JJ), (Fair, NNP), (enough, RB), (B...		1.000	0.125	
5 [(richardebaker, NN), (no, DT), (it, PRP), (is...		1.500	1.125	
6 [(Fuck, NNP), (this, DT), (economy, NN), (I, P...		0.000	0.875	
7 [(Jquery, NN), (is, VBZ), (my, PRP\$), (new, JJ...		1.500	0.125	
8 [(Loves, NNS), (twitter, NN)]		0.625	0.000	
9 [(how, WRB), (can, MD), (you, PRP), (not, RB),...		0.875	0.625	
10 [(Check, VB), (this, DT), (video, NN), (out, R...		0.500	0.000	
11 [(Karloli, NNP), (I, PRP), (firmly, RB), (belie...		1.875	0.500	
12 [(House, NNP), (Correspondents, NNP), (dinner,...		0.875	0.375	
13 [(Watchin, NNP), (Espn, NNP), (Jus, NNP), (see...		0.625	0.375	
14 [(dear, JJ), (nike, IN), (stop, NN), (with, IN...		0.875	0.375	
15 [(best, JJS), (athlete, NN), (of, IN), (our, P...		1.125	0.625	
16 [(I, PRP), (was, VBD), (talking, VBG), (to, TO...		0.750	1.875	
17 [(i, NN), (love, VBP), (lebron, NN)]		0.500	0.000	
18 [(ludajoice, NN), (Lebron, NNP), (is, VBZ), (a...		0.500	0.625	
19 [(Pmillzz, NNP), (lebron, NN), (IS, VBZ), (THE...		0.250	0.125	

#### Note\_globale

0	1
1	1
2	1
3	0
4	1
5	1
6	-1
7	1
8	1
9	1
10	1
11	1
12	1

13	1
14	1
15	1
16	-1
17	1
18	-1
19	1

None

On récupère les notes des tweets: on donne 1 pour 4, 0 pour 2 et -1 pour 0

```
In [37]: df1['Notes_tweet'] = df1['Notes_tweet'].replace(0,-1).replace(2,0).replace(
        print(display(df1.iloc[0:20]))
```

	Notes_tweet	Index	Date	Subject	Author \
0	1	3	Mon May 11 03:17:40 UTC 2009	kindle2	tpryan
1	1	4	Mon May 11 03:18:03 UTC 2009	kindle2	vcu451
2	1	5	Mon May 11 03:18:54 UTC 2009	kindle2	chadfu
3	1	6	Mon May 11 03:19:04 UTC 2009	kindle2	SIX15
4	1	7	Mon May 11 03:21:41 UTC 2009	kindle2	yamarama
5	1	8	Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme
6	-1	9	Mon May 11 03:22:30 UTC 2009	aig	Seth937
7	1	10	Mon May 11 03:26:10 UTC 2009	jquery	dcostalis
8	1	11	Mon May 11 03:27:15 UTC 2009	twitter	PJ_King
9	1	12	Mon May 11 03:29:20 UTC 2009	obama	mandanicole
10	0	13	Mon May 11 03:32:42 UTC 2009	obama	jpeb
11	-1	14	Mon May 11 03:32:48 UTC 2009	obama	kylesellers
12	1	15	Mon May 11 03:33:38 UTC 2009	obama	theviewfans
13	1	16	Mon May 11 05:05:58 UTC 2009	nike	MumsFP
14	-1	17	Mon May 11 05:06:22 UTC 2009	nike	vincentx24x
15	1	18	Mon May 11 05:20:15 UTC 2009	lebron	cameronwylie
16	-1	19	Mon May 11 05:20:28 UTC 2009	lebron	luv8242
17	1	20	Mon May 11 05:21:04 UTC 2009	lebron	mtgillikin
18	-1	21	Mon May 11 05:21:37 UTC 2009	lebron	ursecretdezire
19	1	22	Mon May 11 05:21:45 UTC 2009	lebron	Native_01

	Tweet \
0	@stellargirl I loooooooooovvvvvveee my Kindle2. ...
1	Reading my kindle2... Love it... Lee childs i...
2	Ok, first assesment of the #kindle2 ...it fuck...
3	@kenburbarry You'll love your Kindle2. I've had...
4	@mikefish Fair enough. But i have the Kindle2...
5	@richardebaker no. it is too big. I'm quite ha...
6	Fuck this economy. I hate aig and their non lo...
7	Jquery is my new best friend.
8	Loves twitter

9 how can you not love Obama? he makes jokes abo...  
 10 Check this video out -- President Obama at the...  
 11 @Karloli I firmly believe that Obama/Pelosi hav...  
 12 House Correspondents dinner was last night who...  
 13 Watchin Espn..Jus seen this new Nike Commerica...  
 14 dear nike, stop with the flywire. that shit is...  
 15 #lebron best athlete of our generation, if not...  
 16 I was talking to this guy last night and he wa...  
 17 i love lebron. <http://bit.ly/PdHur>  
 18 @ludajuice Lebron is a Beast, but I'm still ch...  
 19 @Pmillzz lebron IS THE BOSS

clean\_tweet \

0 stellargirl I love my Kindle2 Not that the DX ...  
 1 Reading my kindle2 Love it Lee childs is good ...  
 2 Ok first assesment of the it fucking rocks  
 3 kenburbary You ll love your Kindle2 I ve had m...  
 4 mikefish Fair enough But i have the Kindle2 an...  
 5 richardebaker no it is too big I m quite happy...  
 6 Fuck this economy I hate aig and their non loa...  
 7 JQuery is my new best friend  
 8 Loves twitter  
 9 how can you not love Obama he makes jokes abou...  
 10 Check this video out -- President Obama at the...  
 11 Karoli I firmly believe that Obama/Pelosi have...  
 12 House Correspondents dinner was last night who...  
 13 Watchin Espn Jus seen this new Nike Commerical...  
 14 dear nike stop with the flywire that shit is a...  
 15 best athlete of our generation if not all time...  
 16 I was talking to this guy last night and he wa...  
 17 i love lebron  
 18 ludajuice Lebron is a Beast but I m still chee...  
 19 Pmillzz lebron IS THE BOSS

tokens \

0 [stellargirl, I, love, my, Kindle2, Not, that,...  
 1 [Reading, my, kindle2, Love, it, Lee, childs, ...  
 2 [Ok, first, assesment, of, the, it, fucking, r...  
 3 [kenburbary, You, ll, love, your, Kindle2, I, ...  
 4 [mikefish, Fair, enough, But, i, have, the, Ki...  
 5 [richardebaker, no, it, is, too, big, I, m, qu...  
 6 [Fuck, this, economy, I, hate, aig, and, their...  
 7 [Jquery, is, my, new, best, friend]  
 8 [Loves, twitter]  
 9 [how, can, you, not, love, Obama, he, makes, j...  
 10 [Check, this, video, out, --, President, Obama...  
 11 [Karloli, I, firmly, believe, that, Obama/Pelos...  
 12 [House, Correspondents, dinner, was, last, nig...

13 [Watchin, Espn, Jus, seen, this, new, Nike, Co...  
14 [dear, nike, stop, with, the, flywire, that, s...  
15 [best, athlete, of, our, generation, if, not, ...  
16 [I, was, talking, to, this, guy, last, night, ...  
17 [i, love, lebron]  
18 [ludajoice, Lebron, is, a, Beast, but, I, m, s...  
19 [Pmillzz, lebron, IS, THE, BOSS]

	POS	Score_Pos	Score_Neg	\
0 [(stellargirl, NN), (I, PRP), (love, VBP), (my...		1.250	1.000	
1 [(Reading, VBG), (my, PRP\$), (kindle2, NN), (L...		1.750	0.125	
2 [(Ok, NNP), (first, JJ), (assesment, NN), (of,...		0.375	0.125	
3 [(kenburbary, NN), (You, PRP), (ll, VBP), (lov...		1.875	1.875	
4 [(mikefish, JJ), (Fair, NNP), (enough, RB), (B...		1.000	0.125	
5 [(richardebaker, NN), (no, DT), (it, PRP), (is...		1.500	1.125	
6 [(Fuck, NNP), (this, DT), (economy, NN), (I, P...		0.000	0.875	
7 [(Jquery, NN), (is, VBZ), (my, PRP\$), (new, JJ...		1.500	0.125	
8 [(Loves, NNS), (twitter, NN)]		0.625	0.000	
9 [(how, WRB), (can, MD), (you, PRP), (not, RB),...		0.875	0.625	
10 [(Check, VB), (this, DT), (video, NN), (out, R...		0.500	0.000	
11 [(Karloli, NNP), (I, PRP), (firmly, RB), (belie...		1.875	0.500	
12 [(House, NNP), (Correspondents, NNP), (dinner,...		0.875	0.375	
13 [(Watchin, NNP), (Espn, NNP), (Jus, NNP), (see...		0.625	0.375	
14 [(dear, JJ), (nike, IN), (stop, NN), (with, IN...		0.875	0.375	
15 [(best, JJS), (athlete, NN), (of, IN), (our, P...		1.125	0.625	
16 [(I, PRP), (was, VBD), (talking, VBG), (to, TO...		0.750	1.875	
17 [(i, NN), (love, VBP), (lebron, NN)]		0.500	0.000	
18 [(ludajoice, NN), (Lebron, NNP), (is, VBZ), (a...		0.500	0.625	
19 [(Pmillzz, NNP), (lebron, NN), (IS, VBZ), (THE...		0.250	0.125	

#### Note\_globale

0	1
1	1
2	1
3	0
4	1
5	1
6	-1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	-1

```

17         1
18        -1
19         1

```

None

```

In [38]: n_positif = sum(df1['Notes_tweet']==1)
        n_neutre = sum(df1['Notes_tweet']==0)
        n_negatif = sum(df1['Notes_tweet']==-1)
        print('le nombre de tweets positifs vrais est: ', n_positif)
        print(' ')
        print('le nombre de tweets neutres vrais est: ', n_neutre)
        print(' ')
        print('le nombre de tweets négatifs vrais est: ', n_negatif)

```

le nombre de tweets positifs vrais est: 182

le nombre de tweets neutres vrais est: 139

le nombre de tweets négatifs vrais est: 177

On teste le score de prédiction:

```

In [39]: y_true = np.array(df1['Notes_tweet'])
        y_pred = np.array(df1['Note_globale'])
        print(classification_report(y_true,y_pred))

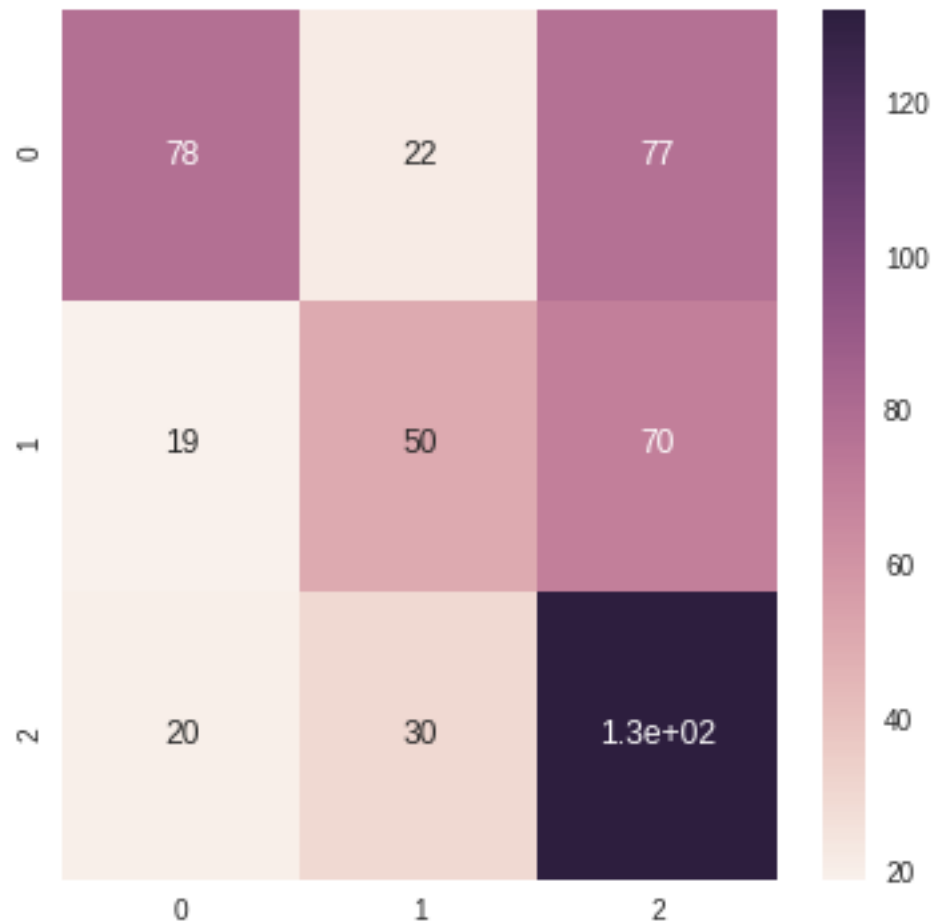
```

	precision	recall	f1-score	support
-1	0.67	0.44	0.53	177
0	0.49	0.36	0.41	139
1	0.47	0.73	0.57	182
avg / total	0.55	0.52	0.51	498

```

In [40]: conf_matrix_meth01 = confusion_matrix(y_true, y_pred)
        fig = plt.figure(figsize=(6,6))
        ax = sns.heatmap(conf_matrix_meth01, annot=True)

```



```
In [41]: list_tweet_pos_correct = [y_pred[i] for i in range(len(y_pred)) if (y_pred[i] == 0)]
print('le nombre de tweets positifs prédits correctement: ', len(list_tweet_pos_correct))
```

le nombre de tweets positifs prédits correctement: 132

```
In [42]: list_tweet_neg_correct = [y_pred[i] for i in range(len(y_pred)) if (y_pred[i] == 1)]
print('le nombre de tweets négatifs prédits correctement: ', len(list_tweet_neg_correct))
```

le nombre de tweets négatifs prédits correctement: 78

```
In [43]: print('le nombre de tweets mal prédits est:', N - sum(y_pred == y_true))
```

le nombre de tweets mal prédits est: 238

### 1.3 1.3 Algorithme de détection v2 : gestion de la négation et des modifieurs

```
In [44]: f = open('./BoosterWordList.txt')
        boosterlines = f.readlines()
        f.close()

In [45]: list_booster = [x.replace('\t',' ').replace('\n','') for x in boosterlines]
        booster = [l.split()[0] for l in list_booster if l]
        score_booster = [' '.join(l.split()[1:]) for l in list_booster if l]
        dico_booster = {}
        for i,j in zip(booster,score_booster):
            dico_booster[i] = j
        print(dico_booster)

{'absolutely': '1', 'overwhelmingly': '2', 'fuckin': '2', 'definitely': '1', 'very': '1'}

In [46]: f = open('./NegatingWordList.txt')
        negatinglines = f.readlines()
        f.close()

In [47]: list_negating = [x.replace('\n','') for x in negatinglines]
        print(list_negating)

["aren't", 'arent', "ain't", "can't", 'cannot', 'cant', "don't", 'dont', "isn't", 'isnt']

In [48]: def get_score_with_neg_mod(tweet):
        new_score_pos = []
        new_score_neg = []
        n = len(tweet)
        if n > 1:
            for i in range (1,n):
                score_pos, score_neg =0,0
                w,t = tweet[i]
                w_1 = tweet[i-1][0]
                tweet_senti = list(swn.senti_synsets(w,t))
                if tweet_senti:
                    senti = list(swn.senti_synsets(w,t))[0]
                    if w_1 in booster:
                        score_pos = 2*senti.pos_score
                        score_neg = 2*senti.neg_score
                    elif w_1 in list_negating:
                        score_pos = senti.neg_score
                        score_neg = senti.pos_score
                    else:
                        score_pos = senti.pos_score
                        score_neg = senti.neg_score
                new_score_pos.append(score_pos)
                new_score_neg.append(score_neg)
        return new_score_pos, new_score_neg, sum(new_score_pos), sum(new_score_neg)
```



```

In [49]: corr_pos_score1 = []
        corr_neg_score1 = []
        corr_pos_glob_score1 = []
        corr_neg_glob_score1 = []
        for t in (tweet_sanitize):
            tweet_pos, tweet_neg, tweet_pos_score, tweet_neg_score = get_score_wit
            corr_pos_score1.append(tweet_pos)
            corr_neg_score1.append(tweet_neg)
            corr_pos_glob_score1.append(tweet_pos_score)
            corr_neg_glob_score1.append(tweet_neg_score)
        print('Score positif global des 10 premiers tweets: ', corr_pos_glob_score
        print('Score negatif global des 10 premiers tweets: ', corr_neg_glob_score

```

```

Score positif global des 10 premiers tweets:  [1.25, 1.75, 0.375, 1.875, 1.0, 1.5,
Score negatif global des 10 premiers tweets:  [1.0, 0.125, 0.125, 1.875, 0.125, 1.1

```

```

In [50]: df1['Score_Pos_v2'] = corr_pos_glob_score1
        df1['Score_Neg_v2'] = corr_neg_glob_score1
        note_global1 = []
        for p,n in zip(corr_pos_glob_score1, corr_neg_glob_score1):
            if p > n:
                note_global1.append(1)
            elif p == n:
                note_global1.append(0)
            else:
                note_global1.append(-1)
        df1['Note_globale_v2'] = note_global1
        print(display(df1.iloc[0:20]))

```

	Notes_tweet	Index	Date	Subject	Author \
0	1	3	Mon May 11 03:17:40 UTC 2009	kindle2	tpryan
1	1	4	Mon May 11 03:18:03 UTC 2009	kindle2	vcu451
2	1	5	Mon May 11 03:18:54 UTC 2009	kindle2	chadfu
3	1	6	Mon May 11 03:19:04 UTC 2009	kindle2	SIX15
4	1	7	Mon May 11 03:21:41 UTC 2009	kindle2	yamarama
5	1	8	Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme
6	-1	9	Mon May 11 03:22:30 UTC 2009	aig	Seth937
7	1	10	Mon May 11 03:26:10 UTC 2009	jquery	dcostalis
8	1	11	Mon May 11 03:27:15 UTC 2009	twitter	PJ_King
9	1	12	Mon May 11 03:29:20 UTC 2009	obama	mandanicole
10	0	13	Mon May 11 03:32:42 UTC 2009	obama	jpeb
11	-1	14	Mon May 11 03:32:48 UTC 2009	obama	kylesellers
12	1	15	Mon May 11 03:33:38 UTC 2009	obama	theviewfans
13	1	16	Mon May 11 05:05:58 UTC 2009	nike	MumsFP
14	-1	17	Mon May 11 05:06:22 UTC 2009	nike	vincentx24x
15	1	18	Mon May 11 05:20:15 UTC 2009	lebron	cameronwylie
16	-1	19	Mon May 11 05:20:28 UTC 2009	lebron	luv8242

17	1	20	Mon May 11 05:21:04 UTC 2009	lebron	mtgillikin
18	-1	21	Mon May 11 05:21:37 UTC 2009	lebron	ursecretdezire
19	1	22	Mon May 11 05:21:45 UTC 2009	lebron	Native_01

Tweet \

```

0  @stellargirl I loooooooooovvvvvveee my Kindle2. ...
1  Reading my kindle2... Love it... Lee child's i...
2  Ok, first assesment of the #kindle2 ...it fuck...
3  @kenburbary You'll love your Kindle2. I've had...
4  @mikefish Fair enough. But i have the Kindle2...
5  @richardebaker no. it is too big. I'm quite ha...
6  Fuck this economy. I hate aig and their non lo...
7      JQuery is my new best friend.
8      Loves twitter
9  how can you not love Obama? he makes jokes abo...
10 Check this video out -- President Obama at the...
11 @KaroLi I firmly believe that Obama/Pelosi hav...
12 House Correspondents dinner was last night who...
13 Watchin Espn..Jus seen this new Nike Commerica...
14 dear nike, stop with the flywire. that shit is...
15 #lebron best athlete of our generation, if not...
16 I was talking to this guy last night and he wa...
17      i love lebron. http://bit.ly/PdHur
18 @ludajuice Lebron is a Beast, but I'm still ch...
19      @Pmillzz lebron IS THE BOSS

```

clean\_tweet \

```

0  stellargirl I love my Kindle2 Not that the DX ...
1  Reading my kindle2 Love it Lee child's is good ...
2      Ok first assesment of the it fucking rocks
3  kenburbary You ll love your Kindle2 I ve had m...
4  mikefish Fair enough But i have the Kindle2 an...
5  richardebaker no it is too big I m quite happy...
6  Fuck this economy I hate aig and their non loa...
7      JQuery is my new best friend
8      Loves twitter
9  how can you not love Obama he makes jokes abou...
10 Check this video out -- President Obama at the...
11 Karoli I firmly believe that Obama/Pelosi have...
12 House Correspondents dinner was last night who...
13 Watchin Espn Jus seen this new Nike Commerical...
14 dear nike stop with the flywire that shit is a...
15 best athlete of our generation if not all time...
16 I was talking to this guy last night and he wa...
17      i love lebron
18 ludajuice Lebron is a Beast but I m still chee...
19      Pmillzz lebron IS THE BOSS

```

	tokens \
0	[stellargirl, I, love, my, Kindle2, Not, that,...
1	[Reading, my, kindle2, Love, it, Lee, childs, ...
2	[Ok, first, assesment, of, the, it, fucking, r...
3	[kenburbary, You, ll, love, your, Kindle2, I, ...
4	[mikefish, Fair, enough, But, i, have, the, Ki...
5	[richardebaker, no, it, is, too, big, I, m, qu...
6	[Fuck, this, economy, I, hate, aig, and, their...
7	[Jquery, is, my, new, best, friend]
8	[Loves, twitter]
9	[how, can, you, not, love, Obama, he, makes, j...
10	[Check, this, video, out, --, President, Obama...
11	[Karoli, I, firmly, believe, that, Obama/Pelos...
12	[House, Correspondents, dinner, was, last, nig...
13	[Watchin, Espn, Jus, seen, this, new, Nike, Co...
14	[dear, nike, stop, with, the, flywire, that, s...
15	[best, athlete, of, our, generation, if, not, ...
16	[I, was, talking, to, this, guy, last, night, ...
17	[i, love, lebron]
18	[ludaj Juice, Lebron, is, a, Beast, but, I, m, s...
19	[Pmillzz, lebron, IS, THE, BOSS]

	POS	Score_Pos	Score_Neg	\
0	[(stellargirl, NN), (I, PRP), (love, VBP), (my...	1.250	1.000	
1	[(Reading, VBG), (my, PRP\$), (kindle2, NN), (L...	1.750	0.125	
2	[(Ok, NNP), (first, JJ), (assesment, NN), (of,...	0.375	0.125	
3	[(kenburbary, NN), (You, PRP), (ll, VBP), (lov...	1.875	1.875	
4	[(mikefish, JJ), (Fair, NNP), (enough, RB), (B...	1.000	0.125	
5	[(richardebaker, NN), (no, DT), (it, PRP), (is...	1.500	1.125	
6	[(Fuck, NNP), (this, DT), (economy, NN), (I, P...	0.000	0.875	
7	[(Jquery, NN), (is, VBZ), (my, PRP\$), (new, JJ...	1.500	0.125	
8	[(Loves, NNS), (twitter, NN)]	0.625	0.000	
9	[(how, WRB), (can, MD), (you, PRP), (not, RB),...	0.875	0.625	
10	[(Check, VB), (this, DT), (video, NN), (out, R...	0.500	0.000	
11	[(Karoli, NNP), (I, PRP), (firmly, RB), (belie...	1.875	0.500	
12	[(House, NNP), (Correspondents, NNP), (dinner,...	0.875	0.375	
13	[(Watchin, NNP), (Espn, NNP), (Jus, NNP), (see...	0.625	0.375	
14	[(dear, JJ), (nike, IN), (stop, NN), (with, IN...	0.875	0.375	
15	[(best, JJS), (athlete, NN), (of, IN), (our, P...	1.125	0.625	
16	[(I, PRP), (was, VBD), (talking, VBG), (to, TO...	0.750	1.875	
17	[(i, NN), (love, VBP), (lebron, NN)]	0.500	0.000	
18	[(ludaj Juice, NN), (Lebron, NNP), (is, VBZ), (a...	0.500	0.625	
19	[(Pmillzz, NNP), (lebron, NN), (IS, VBZ), (THE...	0.250	0.125	

	Note_globale	Score_Pos_v2	Score_Neg_v2	Note_globale_v2
0	1	1.250	1.000	1
1	1	1.750	0.125	1
2	1	0.375	0.125	1

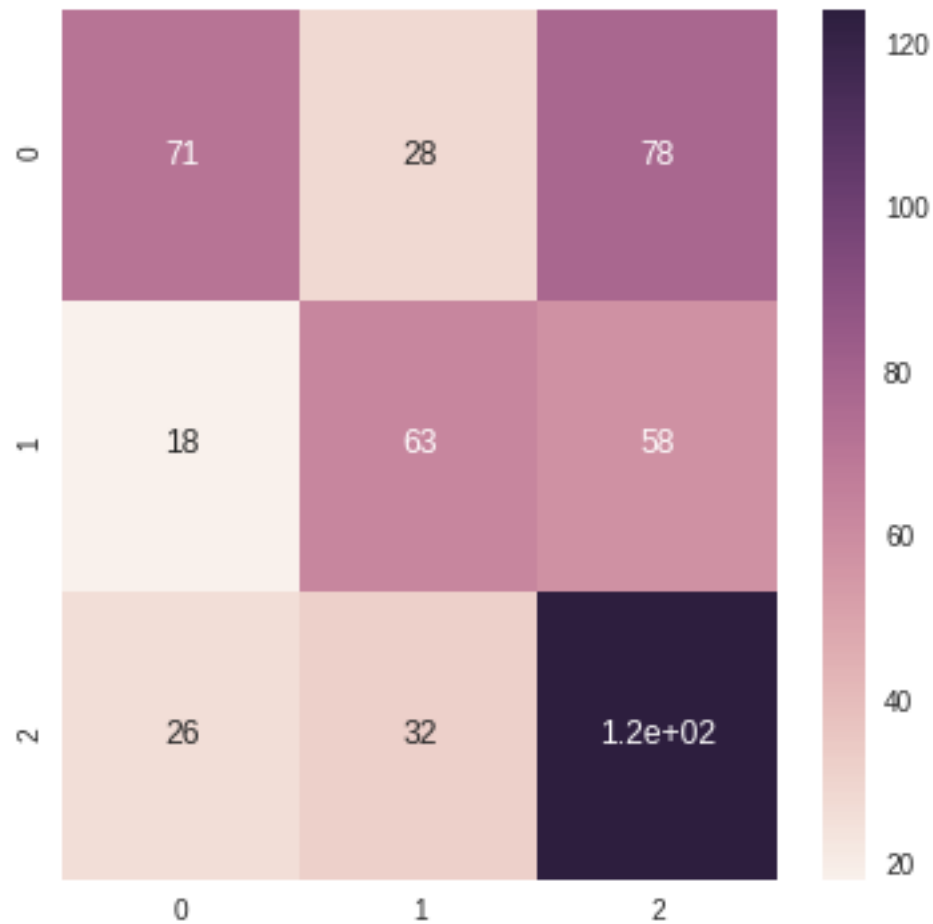
3	0	1.875	1.875	0
4	1	1.000	0.125	1
5	1	1.500	1.125	1
6	-1	0.000	0.875	-1
7	1	1.500	0.125	1
8	1	0.000	0.000	0
9	1	0.375	0.500	-1
10	1	0.000	0.000	0
11	1	1.875	0.500	1
12	1	0.875	0.375	1
13	1	0.625	0.375	1
14	1	0.875	0.375	1
15	1	0.375	0.625	-1
16	-1	0.500	1.750	-1
17	1	0.500	0.000	1
18	-1	0.500	0.625	-1
19	1	0.250	0.125	1

None

```
In [51]: y_pred1 = np.array(df1['Note_globale_v2'])
         print(classification_report(y_true,y_pred1))
```

	precision	recall	f1-score	support
-1	0.62	0.40	0.49	177
0	0.51	0.45	0.48	139
1	0.48	0.68	0.56	182
avg / total	0.54	0.52	0.51	498

```
In [52]: conf_matrix_meth02 = confusion_matrix(y_true, y_pred1)
         fig = plt.figure(figsize=(6,6))
         ax = sns.heatmap(conf_matrix_meth02, annot=True)
```



```
In [53]: list_tweet_pos_correct2 = [y_pred1[i] for i in range(len(y_pred1)) if (y_p
print('le nombre de tweets positifs prédits correctement est de: ', len(list_tweet_pos_correct2))
```

le nombre de tweets positifs prédits correctement est de: 124

```
In [54]: list_tweet_neg_correct2 = [y_pred1[i] for i in range(len(y_pred1)) if (y_p
print('le nombre de tweets negatifs prédits correctement est de: ', len(list_tweet_neg_correct2))
```

le nombre de tweets negatifs prédits correctement est de: 71

Le nombre de tweets positifs prédits correctement avec cette méthode est de 124

```
In [55]: print('le nombre de tweets mal prédits est:', N - sum(y_pred1==y_true))
```

le nombre de tweets mal prédits est: 240

Avec cette méthode on gère moins bien la prédiction des tweets positifs

```
In [56]: df3 = df1[['Notes_tweet', 'Index', 'Date', 'Subject', 'Author', 'Tweet']]

In [57]: f = open('./EmoticonLookupTable.txt')
emoticonlines = f.readlines()
f.close()

In [58]: list_emoticon = [x.replace('\t', 'MM').replace('\n', '') for x in emoticonlines]
emoticon = []
value_emoticon = []
list_emoticon = [l.split('MM') for l in list_emoticon if l]
for e,t in list_emoticon:
    emoticon.append(e)
    value_emoticon.append(int(t))

dico_emoticon = {}
for i,j in zip(emoticon,value_emoticon):
    dico_emoticon[i] = j
print(dico_emoticon)

{'<3': 1, ':-D': 1, '--^--@': 1, ':-*': 1, ':[': -1, 'XP': 1, '>:O': -1, '(^.^)': 1}

In [59]: list_tweet2 = df3['Tweet']
n_url2, n_hashtag2, n_destinataire2, n_retweet2, dico_tweet2, clean_tweet2 = list_tweet2
print(clean_tweet2)

by's been targeted down the mouse up & with earlly by kindle had no more after the 0X like spok, none? dale 2 to

In [60]: tweet_no_emoji = []
table_index = []
score_initial = []
for ind, tweet in zip(df3['Index'],clean_tweet2):
    for k,v in dico_emoticon.items():
        if k in tweet:
            t = tweet.replace(k, ' ')
            tweet_no_emoji.append(t)
            table_index.append(ind)
            score_initial.append(v)

print(tweet_no_emoji)
print(' ')
print(score_initial)
print(' ')
print(table_index)

["kenburbary You'll love your Kindle2. I've had mine for a few months and never lo
```

```
[1, 1, 1, 1, 1, 1, 0, -1, 1, 1, -1, -1, 1, -1, -1, -1, -1, -1, 1, -1, -1, -1, 1, 0,
[6, 7, 121, 131, 131, 132, 153, 177, 178, 195, 196, 213, 217, 219, 220, 222, 228, 4
```

```
In [61]: total_tweet = []
        total_score_init = []
        i = 0
        for ind, tweet in zip(df3['Index'], clean_tweet2):
            if ind not in table_index:
                total_tweet.append(tweet)
                total_score_init.append(0)
            else:
                total_tweet.append(tweet_no_emoji[i])
                total_score_init.append(score_initial[i])
                i += 1
```

```
In [62]: df3['Tweet_no_emoji'] = total_tweet
        df3['score_no_emoji'] = total_score_init
        print(display(df3.iloc[:10]))
```

/home/audrey/anaconda3/lib/python3.4/site-packages/ipykernel/\_\_main\_\_.py:1: Setting a value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/>  
if \_\_name\_\_ == '\_\_main\_\_':  
/home/audrey/anaconda3/lib/python3.4/site-packages/ipykernel/\_\_main\_\_.py:2: Setting a value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/>  
from ipykernel import kernelapp as app

	Notes_tweet	Index	Date	Subject	Author \
0	1	3 Mon May 11 03:17:40 UTC 2009	kindle2	tpryan	
1	1	4 Mon May 11 03:18:03 UTC 2009	kindle2	vcu451	
2	1	5 Mon May 11 03:18:54 UTC 2009	kindle2	chadfu	
3	1	6 Mon May 11 03:19:04 UTC 2009	kindle2	SIX15	
4	1	7 Mon May 11 03:21:41 UTC 2009	kindle2	yamarama	
5	1	8 Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme	
6	-1	9 Mon May 11 03:22:30 UTC 2009	aig	Seth937	
7	1	10 Mon May 11 03:26:10 UTC 2009	jquery	dcostalis	
8	1	11 Mon May 11 03:27:15 UTC 2009	twitter	PJ_King	
9	1	12 Mon May 11 03:29:20 UTC 2009	obama	mandanicole	

```

                                Tweet \
0 @stellargirl I loooooooooovvvvvveee my Kindle2. ...
```

```

1 Reading my kindle2... Love it... Lee childs i...
2 Ok, first assesment of the #kindle2 ...it fuck...
3 @kenburbary You'll love your Kindle2. I've had...
4 @mikefish Fair enough. But i have the Kindle2...
5 @richardebaker no. it is too big. I'm quite ha...
6 Fuck this economy. I hate aig and their non lo...
7 Jquery is my new best friend.
8 Loves twitter
9 how can you not love Obama? he makes jokes abo...

```

	Tweet_no_emoji	score_no_emoji
0 stellargirl I loooooooooovvvvvveee my Kindle2. N...		0
1 Reading my kindle2... Love it... Lee childs i...		0
2 Ok, first assesment of the ...it fucking rock...		0
3 kenburbary You'll love your Kindle2. I've had ...		1
4 mikefish Fair enough. But i have the Kindle2 ...		1
5 richardebaker no. it is too big. I'm quite hap...		0
6 Fuck this economy. I hate aig and their non lo...		0
7 Jquery is my new best friend.		0
8 Loves twitter		0
9 how can you not love Obama? he makes jokes abo...		0

None

```

In [63]: _, _, _, _, dico_tweet3, clean_tweet3 = apply_clean_listv1(clean_tweet2, h
print(clean_tweet3)

```

\\s10'got'Ghómé's'fó'mé'c'í'h'f'í'á'c'ó'm'b'ó'u's'm'ó'y'e't'h'á'm'í'n'y'h'e'g't'N'o't'h'a'd'o'n'l'y' D'x'is'c'o'k'e'y'á'r'n'e'á'2'd'á'

```

In [64]: wordsnoslang2 = clean_from_dico(clean_tweet3, dico_slang, lower=False)
print(wordsnoslang2)

```

```

[['stellargirl', 'I', 'love', 'my', 'Kindle2', 'Not', 'that', 'the', 'DX', 'is', 'c

```



A value is trying to be set on a copy of a slice from a DataFrame.  
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/>  
from ipykernel import kernelapp as app

	Notes_tweet	Index	Date	Subject	Author \
0	1	3	Mon May 11 03:17:40 UTC 2009	kindle2	tpryan
1	1	4	Mon May 11 03:18:03 UTC 2009	kindle2	vcu451
2	1	5	Mon May 11 03:18:54 UTC 2009	kindle2	chadfu
3	1	6	Mon May 11 03:19:04 UTC 2009	kindle2	SIX15
4	1	7	Mon May 11 03:21:41 UTC 2009	kindle2	yamarama
5	1	8	Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme
6	-1	9	Mon May 11 03:22:30 UTC 2009	aig	Seth937
7	1	10	Mon May 11 03:26:10 UTC 2009	jquery	dcostalis
8	1	11	Mon May 11 03:27:15 UTC 2009	twitter	PJ_King
9	1	12	Mon May 11 03:29:20 UTC 2009	obama	mandanicole
10	0	13	Mon May 11 03:32:42 UTC 2009	obama	jpeb
11	-1	14	Mon May 11 03:32:48 UTC 2009	obama	kylesellers
12	1	15	Mon May 11 03:33:38 UTC 2009	obama	theviewfans
13	1	16	Mon May 11 05:05:58 UTC 2009	nike	MumsFP
14	-1	17	Mon May 11 05:06:22 UTC 2009	nike	vincentx24x
15	1	18	Mon May 11 05:20:15 UTC 2009	lebron	cameronwylie
16	-1	19	Mon May 11 05:20:28 UTC 2009	lebron	luv8242
17	1	20	Mon May 11 05:21:04 UTC 2009	lebron	mtgillikin
18	-1	21	Mon May 11 05:21:37 UTC 2009	lebron	ursecretdezire
19	1	22	Mon May 11 05:21:45 UTC 2009	lebron	Native_01

	Tweet \
0	@stellargirl I loooooooooovvvvvveee my Kindle2. ...
1	Reading my kindle2... Love it... Lee childs i...
2	Ok, first assesment of the #kindle2 ...it fuck...
3	@kenburbary You'll love your Kindle2. I've had...
4	@mikefish Fair enough. But i have the Kindle2...
5	@richardebaker no. it is too big. I'm quite ha...
6	Fuck this economy. I hate aig and their non lo...
7	jQuery is my new best friend.
8	Loves twitter
9	how can you not love Obama? he makes jokes abo...
10	Check this video out -- President Obama at the...
11	@Karloli I firmly believe that Obama/Pelosi hav...
12	House Correspondents dinner was last night who...
13	Watchin Espn..Jus seen this new Nike Commerica...
14	dear nike, stop with the flywire. that shit is...
15	#lebron best athlete of our generation, if not...
16	I was talking to this guy last night and he wa...
17	i love lebron. <a href="http://bit.ly/PdHur">http://bit.ly/PdHur</a>

18 @ludajoice Lebron is a Beast, but I'm still ch...  
 19 @Pmillzz lebron IS THE BOSS

	Tweet_no_emoji	score_no_emoji	\
0	stellargirl I loooooooooovvvvvveee my Kindle2. N...	0	
1	Reading my kindle2... Love it... Lee childs i...	0	
2	Ok, first assesment of the ...it fucking rock...	0	
3	kenburbary You'll love your Kindle2. I've had ...	1	
4	mikefish Fair enough. But i have the Kindle2 ...	1	
5	richardebaker no. it is too big. I'm quite hap...	0	
6	Fuck this economy. I hate aig and their non lo...	0	
7	Jquery is my new best friend.	0	
8	Loves twitter	0	
9	how can you not love Obama? he makes jokes abo...	0	
10	Check this video out -- President Obama at the...	0	
11	Karoli I firmly believe that Obama/Pelosi have...	0	
12	House Correspondents dinner was last night who...	0	
13	Watchin Espn..Jus seen this new Nike Commerica...	0	
14	dear nike, stop with the flywire. that shit is...	0	
15	best athlete of our generation, if not all ti...	0	
16	I was talking to this guy last night and he wa...	0	
17	i love lebron.	0	
18	ludajoice Lebron is a Beast, but I'm still che...	0	
19	Pmillzz lebron IS THE BOSS	0	

	clean_tweet	\
0	stellargirl I love my Kindle2 Not that the DX ...	
1	Reading my kindle2 Love it Lee childs is good ...	
2	Ok first assesment of the it fucking rocks	
3	kenburbary You ll love your Kindle2 I ve had m...	
4	mikefish Fair enough But i have the Kindle2 an...	
5	richardebaker no it is too big I m quite happy...	
6	Fuck this economy I hate aig and their non loa...	
7	Jquery is my new best friend	
8	Loves twitter	
9	how can you not love Obama he makes jokes abou...	
10	Check this video out -- President Obama at the...	
11	Karoli I firmly believe that Obama/Pelosi have...	
12	House Correspondents dinner was last night who...	
13	Watchin Espn Jus seen this new Nike Commerical...	
14	dear nike stop with the flywire that shit is a...	
15	best athlete of our generation if not all time...	
16	I was talking to this guy last night and he wa...	
17	i love lebron	
18	ludajoice Lebron is a Beast but I m still chee...	
19	Pmillzz lebron IS THE BOSS	

tokens

```

0  [stellargirl, I, love, my, Kindle2, Not, that,...
1  [Reading, my, kindle2, Love, it, Lee, childs, ...
2  [Ok, first, assesment, of, the, it, fucking, r...
3  [kenburbary, You, ll, love, your, Kindle2, I, ...
4  [mikefish, Fair, enough, But, i, have, the, Ki...
5  [richardebaker, no, it, is, too, big, I, m, qu...
6  [Fuck, this, economy, I, hate, aig, and, their...
7      [Jquery, is, my, new, best, friend]
8      [Loves, twitter]
9  [how, can, you, not, love, Obama, he, makes, j...
10 [Check, this, video, out, --, President, Obama...
11 [Karoli, I, firmly, believe, that, Obama/Pelos...
12 [House, Correspondents, dinner, was, last, nig...
13 [Watchin, Espn, Jus, seen, this, new, Nike, Co...
14 [dear, nike, stop, with, the, flywire, that, s...
15 [best, athlete, of, our, generation, if, not, ...
16 [I, was, talking, to, this, guy, last, night, ...
17      [i, love, lebron]
18 [ludaj Juice, Lebron, is, a, Beast, but, I, m, s...
19      [Pmillzz, lebron, IS, THE, BOSS]

```

None

```

In [66]: tag_list_tot2 = getPOS(wordsnoslang2)
         df3['POS'] = tag_list_tot2
         print(display(df3.iloc[0:20]))

```

/home/audrey/anaconda3/lib/python3.4/site-packages/ipykernel/\_\_main\_\_.py:2: Setting a value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/>  
from ipykernel import kernelapp as app

	Notes_tweet	Index	Date	Subject	Author \
0	1	3	Mon May 11 03:17:40 UTC 2009	kindle2	tpryan
1	1	4	Mon May 11 03:18:03 UTC 2009	kindle2	vcu451
2	1	5	Mon May 11 03:18:54 UTC 2009	kindle2	chadfu
3	1	6	Mon May 11 03:19:04 UTC 2009	kindle2	SIX15
4	1	7	Mon May 11 03:21:41 UTC 2009	kindle2	yamarama
5	1	8	Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme
6	-1	9	Mon May 11 03:22:30 UTC 2009	aig	Seth937
7	1	10	Mon May 11 03:26:10 UTC 2009	jquery	dcostalis
8	1	11	Mon May 11 03:27:15 UTC 2009	twitter	PJ_King
9	1	12	Mon May 11 03:29:20 UTC 2009	obama	mandanicole
10	0	13	Mon May 11 03:32:42 UTC 2009	obama	jpeb

11	-1	14	Mon May 11 03:32:48 UTC 2009	obama	kylesellers
12	1	15	Mon May 11 03:33:38 UTC 2009	obama	theviewfans
13	1	16	Mon May 11 05:05:58 UTC 2009	nike	MumsFP
14	-1	17	Mon May 11 05:06:22 UTC 2009	nike	vincentx24x
15	1	18	Mon May 11 05:20:15 UTC 2009	lebron	cameronwylie
16	-1	19	Mon May 11 05:20:28 UTC 2009	lebron	luv8242
17	1	20	Mon May 11 05:21:04 UTC 2009	lebron	mtgillikin
18	-1	21	Mon May 11 05:21:37 UTC 2009	lebron	ursecretdezire
19	1	22	Mon May 11 05:21:45 UTC 2009	lebron	Native_01

Tweet \

```

0  @stellargirl I loooooooooovvvvvveee my Kindle2. ...
1  Reading my kindle2... Love it... Lee childs i...
2  Ok, first assesment of the #kindle2 ...it fuck...
3  @kenburbary You'll love your Kindle2. I've had...
4  @mikefish Fair enough. But i have the Kindle2...
5  @richardebaker no. it is too big. I'm quite ha...
6  Fuck this economy. I hate aig and their non lo...
7      JQuery is my new best friend.
8      Loves twitter
9  how can you not love Obama? he makes jokes abo...
10 Check this video out -- President Obama at the...
11 @KaroIi I firmly believe that Obama/Pelosi hav...
12 House Correspondents dinner was last night who...
13 Watchin Espn..Jus seen this new Nike Commerica...
14 dear nike, stop with the flywire. that shit is...
15 #lebron best athlete of our generation, if not...
16 I was talking to this guy last night and he wa...
17      i love lebron. http://bit.ly/PdHur
18 @ludaj Juice LeBron is a Beast, but I'm still ch...
19      @Pmillzz lebron IS THE BOSS

```

Tweet\_no\_emoji score\_no\_emoji \

0	stellargirl I loooooooooovvvvvveee my Kindle2. N...	0
1	Reading my kindle2... Love it... Lee childs i...	0
2	Ok, first assesment of the ...it fucking rock...	0
3	kenburbary You'll love your Kindle2. I've had ...	1
4	mikefish Fair enough. But i have the Kindle2 ...	1
5	richardebaker no. it is too big. I'm quite hap...	0
6	Fuck this economy. I hate aig and their non lo...	0
7	Jquery is my new best friend.	0
8	Loves twitter	0
9	how can you not love Obama? he makes jokes abo...	0
10	Check this video out -- President Obama at the...	0
11	KaroIi I firmly believe that Obama/Pelosi have...	0
12	House Correspondents dinner was last night who...	0
13	Watchin Espn..Jus seen this new Nike Commerica...	0
14	dear nike, stop with the flywire. that shit is...	0

15	best athlete of our generation, if not all ti...	0
16	I was talking to this guy last night and he wa...	0
17	i love lebron.	0
18	ludajoice Lebron is a Beast, but I'm still che...	0
19	Pmillzz lebron IS THE BOSS	0

	clean_tweet \
0	stellargirl I love my Kindle2 Not that the DX ...
1	Reading my kindle2 Love it Lee childs is good ...
2	Ok first assesment of the it fucking rocks
3	kenburbary You ll love your Kindle2 I ve had m...
4	mikefish Fair enough But i have the Kindle2 an...
5	richardebaker no it is too big I m quite happy...
6	Fuck this economy I hate aig and their non loa...
7	Jquery is my new best friend
8	Loves twitter
9	how can you not love Obama he makes jokes abou...
10	Check this video out -- President Obama at the...
11	Karoli I firmly believe that Obama/Pelosi have...
12	House Correspondents dinner was last night who...
13	Watchin Espn Jus seen this new Nike Commerical...
14	dear nike stop with the flywire that shit is a...
15	best athlete of our generation if not all time...
16	I was talking to this guy last night and he wa...
17	i love lebron
18	ludajoice Lebron is a Beast but I m still chee...
19	Pmillzz lebron IS THE BOSS

	tokens \
0	[stellargirl, I, love, my, Kindle2, Not, that, ...
1	[Reading, my, kindle2, Love, it, Lee, childs, ...
2	[Ok, first, assesment, of, the, it, fucking, r...
3	[kenburbary, You, ll, love, your, Kindle2, I, ...
4	[mikefish, Fair, enough, But, i, have, the, Ki...
5	[richardebaker, no, it, is, too, big, I, m, qu...
6	[Fuck, this, economy, I, hate, aig, and, their...
7	[Jquery, is, my, new, best, friend]
8	[Loves, twitter]
9	[how, can, you, not, love, Obama, he, makes, j...
10	[Check, this, video, out, --, President, Obama...
11	[Karoli, I, firmly, believe, that, Obama/Pelos...
12	[House, Correspondents, dinner, was, last, nig...
13	[Watchin, Espn, Jus, seen, this, new, Nike, Co...
14	[dear, nike, stop, with, the, flywire, that, s...
15	[best, athlete, of, our, generation, if, not, ...
16	[I, was, talking, to, this, guy, last, night, ...
17	[i, love, lebron]
18	[ludajoice, Lebron, is, a, Beast, but, I, m, s...

```
19 [Pmillzz, lebron, IS, THE, BOSS]
```

```
POS
0 [(stellargirl, NN), (I, PRP), (love, VBP), (my...
1 [(Reading, VBG), (my, PRP$), (kindle2, NN), (L...
2 [(Ok, NNP), (first, JJ), (assessment, NN), (of,...
3 [(kenburbary, NN), (You, PRP), (ll, VBP), (lov...
4 [(mikefish, JJ), (Fair, NNP), (enough, RB), (B...
5 [(richardebaker, NN), (no, DT), (it, PRP), (is...
6 [(Fuck, NNP), (this, DT), (economy, NN), (I, P...
7 [(Jquery, NN), (is, VBZ), (my, PRP$), (new, JJ...
8 [(Loves, NNS), (twitter, NN)]
9 [(how, WRB), (can, MD), (you, PRP), (not, RB),...
10 [(Check, VB), (this, DT), (video, NN), (out, R...
11 [(Karloli, NNP), (I, PRP), (firmly, RB), (belie...
12 [(House, NNP), (Correspondents, NNP), (dinner,...
13 [(Watchin, NNP), (Espin, NNP), (Jus, NNP), (see...
14 [(dear, JJ), (nike, IN), (stop, NN), (with, IN...
15 [(best, JJS), (athlete, NN), (of, IN), (our, P...
16 [(I, PRP), (was, VBD), (talking, VBG), (to, TO...
17 [(i, NN), (love, VBP), (lebron, NN)]
18 [(ludajuice, NN), (Lebron, NNP), (is, VBZ), (a...
19 [(Pmillzz, NNP), (lebron, NN), (IS, VBZ), (THE...
```

None

```
In [68]: sum_vb2, count_VB2 = get_verbs(tag_list_tot2)
        print(count_VB2)
        print(' ')
        print('Le nombre total de POS verbes sur l\'ensemble des tweets est: ', su
```

```
Counter({'NNP': 1305, 'NN': 1118, 'IN': 579, 'DT': 435, 'JJ': 434, 'PRP': 373, 'RB': 1170})
```

```
Le nombre total de POS verbes sur l'ensemble des tweets est: 1170
```

Le fait d'enlever les emojis ne changent pas le nombre de verbes détectés (et c'est tant mieux)  
On utilise maintenant sentiword pour donner un score au tweet

```
In [69]: tweet_sanitize2, pos_score2, pos_sum_score2, neg_score2, neg_sum_score2 =
        print(tweet_sanitize2[0])
```

```
[('stellargirl', 'n'), ('love', 'v'), ('kindle2', 'n'), ('not', 'r'), ('dx', 'n'),
```

```
In [70]: corr_pos_score2 = []
        corr_neg_score2 = []
```

```

corr_pos_glob_score2 = []
corr_neg_glob_score2 = []
for t, s in zip(tweet_sanitiz2, total_score_init):
    tweet_pos2, tweet_neg2, _, _ = get_score_with_neg_mod(t)
    if s==1:
        tweet_pos2.extend([1.0])
    elif s==-1:
        tweet_neg2.extend([1.0])
    else:
        pass
corr_pos_score2.append(tweet_pos2)
corr_neg_score2.append(tweet_neg2)

corr_pos_glob_score2 = [sum(corr_pos_score2[i]) for i in range(len(corr_pos_score2))]
corr_neg_glob_score2 = [sum(corr_neg_score2[i]) for i in range(len(corr_neg_score2))]
print('Score positif global des 10 premiers tweets: ', corr_pos_glob_score2)
print('Score negatif global des 10 premiers tweets: ', corr_neg_glob_score2)

```

Score positif global des 10 premiers tweets: [1.25, 1.75, 0.375, 2.875, 2.0, 1.5, 1.0, 1.0, 1.0, 1.0]  
 Score negatif global des 10 premiers tweets: [1.0, 0.125, 0.125, 1.875, 0.125, 1.125, 0.125, 0.125, 0.125, 0.125]

```

In [71]: df3['Score_Pos_v2'] = corr_pos_glob_score2
df3['Score_Neg_v2'] = corr_neg_glob_score2
note_global2 = []
for p,n in zip(corr_pos_glob_score2, corr_neg_glob_score2):
    if p > n:
        note_global2.append(1)
    elif p == n:
        note_global2.append(0)
    else:
        note_global2.append(-1)
df3['Note_globale_v2'] = note_global2
print(display(df3.iloc[0:20]))

```

/home/audrey/anaconda3/lib/python3.4/site-packages/ipykernel/\_\_main\_\_.py:1: Setting a value is trying to be set on a copy of a slice from a DataFrame.  
 Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/10min.html>  
 if \_\_name\_\_ == '\_\_main\_\_':  
 /home/audrey/anaconda3/lib/python3.4/site-packages/ipykernel/\_\_main\_\_.py:2: Setting a value is trying to be set on a copy of a slice from a DataFrame.  
 Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/10min.html>  
 from ipykernel import kernelapp as app  
 /home/audrey/anaconda3/lib/python3.4/site-packages/ipykernel/\_\_main\_\_.py:11: Setting a value is trying to be set on a copy of a slice from a DataFrame.  
 Try using .loc[row\_indexer,col\_indexer] = value instead

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/>

	Notes_tweet	Index	Date	Subject	Author \
0	1	3	Mon May 11 03:17:40 UTC 2009	kindle2	tpryan
1	1	4	Mon May 11 03:18:03 UTC 2009	kindle2	vcu451
2	1	5	Mon May 11 03:18:54 UTC 2009	kindle2	chadfu
3	1	6	Mon May 11 03:19:04 UTC 2009	kindle2	SIX15
4	1	7	Mon May 11 03:21:41 UTC 2009	kindle2	yamarama
5	1	8	Mon May 11 03:22:00 UTC 2009	kindle2	GeorgeVHulme
6	-1	9	Mon May 11 03:22:30 UTC 2009	aig	Seth937
7	1	10	Mon May 11 03:26:10 UTC 2009	jquery	dcostalis
8	1	11	Mon May 11 03:27:15 UTC 2009	twitter	PJ_King
9	1	12	Mon May 11 03:29:20 UTC 2009	obama	mandanicole
10	0	13	Mon May 11 03:32:42 UTC 2009	obama	jpeb
11	-1	14	Mon May 11 03:32:48 UTC 2009	obama	kylesellers
12	1	15	Mon May 11 03:33:38 UTC 2009	obama	theviewfans
13	1	16	Mon May 11 05:05:58 UTC 2009	nike	MumsFP
14	-1	17	Mon May 11 05:06:22 UTC 2009	nike	vincentx24x
15	1	18	Mon May 11 05:20:15 UTC 2009	lebron	cameronwylie
16	-1	19	Mon May 11 05:20:28 UTC 2009	lebron	luv8242
17	1	20	Mon May 11 05:21:04 UTC 2009	lebron	mtgillikin
18	-1	21	Mon May 11 05:21:37 UTC 2009	lebron	ursecretdezire
19	1	22	Mon May 11 05:21:45 UTC 2009	lebron	Native_01

	Tweet \
0	@stellargirl I loooooooooovvvvvveee my Kindle2. ...
1	Reading my kindle2... Love it... Lee childs i...
2	Ok, first assesment of the #kindle2 ...it fuck...
3	@kenburbary You'll love your Kindle2. I've had...
4	@mikefish Fair enough. But i have the Kindle2...
5	@richardebaker no. it is too big. I'm quite ha...
6	Fuck this economy. I hate aig and their non lo...
7	jQuery is my new best friend.
8	Loves twitter
9	how can you not love Obama? he makes jokes abo...
10	Check this video out -- President Obama at the...
11	@KaroIi I firmly believe that Obama/Pelosi hav...
12	House Correspondents dinner was last night who...
13	Watchin Espn..Jus seen this new Nike Commerica...
14	dear nike, stop with the flywire. that shit is...
15	#lebron best athlete of our generation, if not...
16	I was talking to this guy last night and he wa...
17	i love lebron. <a href="http://bit.ly/PdHur">http://bit.ly/PdHur</a>
18	@ludaj Juice LeBron is a Beast, but I'm still ch...



19

@Pmillzz lebron IS THE BOSS

	Tweet_no_emoji	score_no_emoji	\
0	stellargirl I loooooooooovvvvvveee my Kindle2. N...	0	
1	Reading my kindle2... Love it... Lee childs i...	0	
2	Ok, first assesment of the ...it fucking rock...	0	
3	kenburbary You'll love your Kindle2. I've had ...	1	
4	mikefish Fair enough. But i have the Kindle2 ...	1	
5	richardebaker no. it is too big. I'm quite hap...	0	
6	Fuck this economy. I hate aig and their non lo...	0	
7	Jquery is my new best friend.	0	
8	Loves twitter	0	
9	how can you not love Obama? he makes jokes abo...	0	
10	Check this video out -- President Obama at the...	0	
11	Karoli I firmly believe that Obama/Pelosi have...	0	
12	House Correspondents dinner was last night who...	0	
13	Watchin Espn..Jus seen this new Nike Commerica...	0	
14	dear nike, stop with the flywire. that shit is...	0	
15	best athlete of our generation, if not all ti...	0	
16	I was talking to this guy last night and he wa...	0	
17	i love lebron.	0	
18	ludajoice Lebron is a Beast, but I'm still che...	0	
19	Pmillzz lebron IS THE BOSS	0	

	clean_tweet	\
0	stellargirl I love my Kindle2 Not that the DX ...	
1	Reading my kindle2 Love it Lee childs is good ...	
2	Ok first assesment of the it fucking rocks	
3	kenburbary You ll love your Kindle2 I ve had m...	
4	mikefish Fair enough But i have the Kindle2 an...	
5	richardebaker no it is too big I m quite happy...	
6	Fuck this economy I hate aig and their non loa...	
7	Jquery is my new best friend	
8	Loves twitter	
9	how can you not love Obama he makes jokes abou...	
10	Check this video out -- President Obama at the...	
11	Karoli I firmly believe that Obama/Pelosi have...	
12	House Correspondents dinner was last night who...	
13	Watchin Espn Jus seen this new Nike Commerical...	
14	dear nike stop with the flywire that shit is a...	
15	best athlete of our generation if not all time...	
16	I was talking to this guy last night and he wa...	
17	i love lebron	
18	ludajoice Lebron is a Beast but I m still chee...	
19	Pmillzz lebron IS THE BOSS	

	tokens	\
0	[stellargirl, I, love, my, Kindle2, Not, that,...	

1 [Reading, my, kindle2, Love, it, Lee, childs, ...  
 2 [Ok, first, assesment, of, the, it, fucking, r...  
 3 [kenburbary, You, ll, love, your, Kindle2, I, ...  
 4 [mikefish, Fair, enough, But, i, have, the, Ki...  
 5 [richardebaker, no, it, is, too, big, I, m, qu...  
 6 [Fuck, this, economy, I, hate, aig, and, their...  
 7 [Jquery, is, my, new, best, friend]  
 8 [Loves, twitter]  
 9 [how, can, you, not, love, Obama, he, makes, j...  
 10 [Check, this, video, out, --, President, Obama...  
 11 [Karloli, I, firmly, believe, that, Obama/Pelos...  
 12 [House, Correspondents, dinner, was, last, nig...  
 13 [Watchin, Espn, Jus, seen, this, new, Nike, Co...  
 14 [dear, nike, stop, with, the, flywire, that, s...  
 15 [best, athlete, of, our, generation, if, not, ...  
 16 [I, was, talking, to, this, guy, last, night, ...  
 17 [i, love, lebron]  
 18 [ludajuice, Lebron, is, a, Beast, but, I, m, s...  
 19 [Pmillzz, lebron, IS, THE, BOSS]

	POS	Score_Pos_v2	\
0 [(stellargirl, NN), (I, PRP), (love, VBP), (my...		1.250	
1 [(Reading, VBG), (my, PRP\$), (kindle2, NN), (L...		1.750	
2 [(Ok, NNP), (first, JJ), (assesment, NN), (of,...		0.375	
3 [(kenburbary, NN), (You, PRP), (ll, VBP), (lov...		2.875	
4 [(mikefish, JJ), (Fair, NNP), (enough, RB), (B...		2.000	
5 [(richardebaker, NN), (no, DT), (it, PRP), (is...		1.500	
6 [(Fuck, NNP), (this, DT), (economy, NN), (I, P...		0.000	
7 [(Jquery, NN), (is, VBZ), (my, PRP\$), (new, JJ...		1.500	
8 [(Loves, NNS), (twitter, NN)]		0.000	
9 [(how, WRB), (can, MD), (you, PRP), (not, RB),...		0.375	
10 [(Check, VB), (this, DT), (video, NN), (out, R...		0.000	
11 [(Karloli, NNP), (I, PRP), (firmly, RB), (belie...		1.875	
12 [(House, NNP), (Correspondents, NNP), (dinner,...		0.875	
13 [(Watchin, NNP), (Espn, NNP), (Jus, NNP), (see...		0.625	
14 [(dear, JJ), (nike, IN), (stop, NN), (with, IN...		0.875	
15 [(best, JJS), (athlete, NN), (of, IN), (our, P...		0.375	
16 [(I, PRP), (was, VBD), (talking, VBG), (to, TO...		0.500	
17 [(i, NN), (love, VBP), (lebron, NN)]		0.500	
18 [(ludajuice, NN), (Lebron, NNP), (is, VBZ), (a...		0.500	
19 [(Pmillzz, NNP), (lebron, NN), (IS, VBZ), (THE...		0.250	

	Score_Neg_v2	Note_globale_v2
0	1.000	1
1	0.125	1
2	0.125	1
3	1.875	1
4	0.125	1

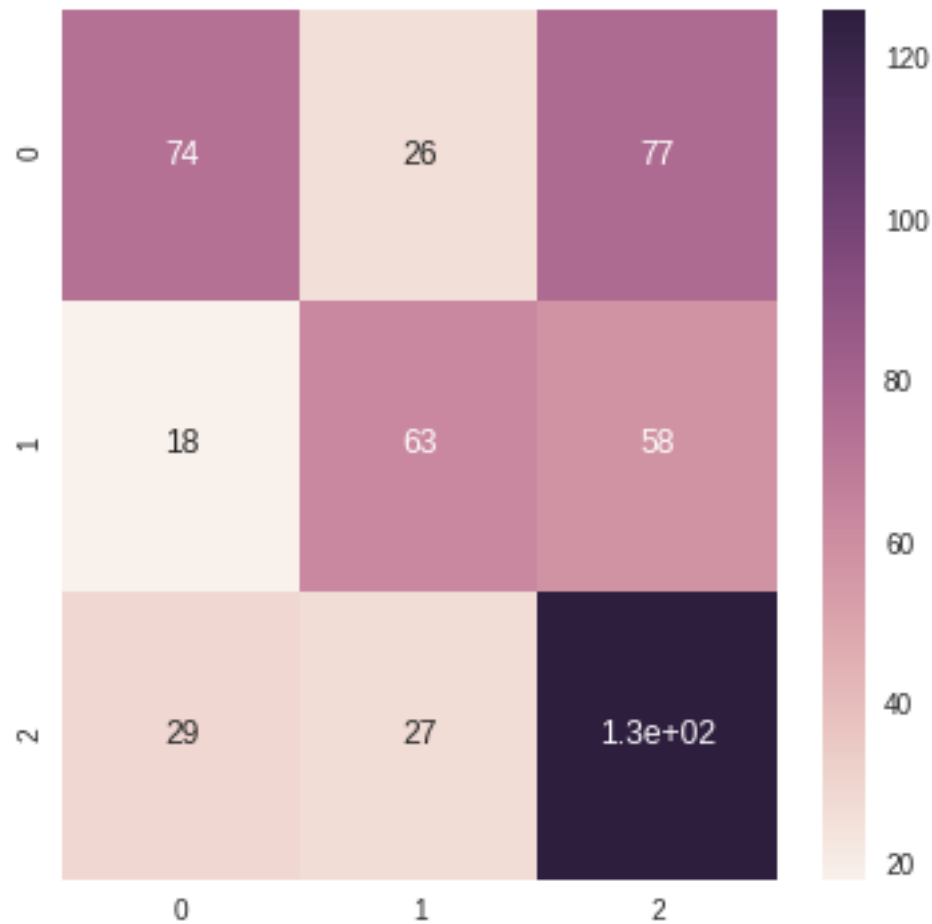
5	1.125	1
6	0.875	-1
7	0.125	1
8	0.000	0
9	0.500	-1
10	0.000	0
11	0.500	1
12	0.375	1
13	0.375	1
14	0.375	1
15	0.625	-1
16	1.750	-1
17	0.000	1
18	0.625	-1
19	0.125	1

None

```
In [72]: y_pred2 = np.array(df3['Note_globale_v2'])
         print(classification_report(y_true,y_pred2))
```

	precision	recall	f1-score	support
-1	0.61	0.42	0.50	177
0	0.54	0.45	0.49	139
1	0.48	0.69	0.57	182
avg / total	0.55	0.53	0.52	498

```
In [73]: conf_matrix_meth02 = confusion_matrix(y_true, y_pred2)
         fig = plt.figure(figsize=(6,6))
         ax = sns.heatmap(conf_matrix_meth02, annot=True)
```



```
In [74]: list_tweet_pos_correct3 = [y_pred2[i] for i in range(len(y_pred2)) if (y_p
print('le nombre de tweets positifs prédits correctement est de : ',len(list_tweet_pos_correct3))
```

le nombre de tweets positifs prédits correctement est de : 126

```
In [75]: print('le nombre de tweets mal prédits est:', N - sum(y_pred2==y_true))
```

le nombre de tweets mal prédits est: 235

## 1.5 1.5 Test d'un autre algorithme

On reprend le travail fait en MDI343 avec l'implémentation du Naive Bayes.

```
In [76]: from sklearn.model_selection import cross_val_score
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.model_selection import GridSearchCV
from sklearn.pipeline import Pipeline
```

```
In [77]: path_stop = './english.stop'
         stop_words = open(path_stop).read().split('\n')
         print(stop_words)
         print('=====')
         print('taille de la liste stop_words:', len(stop_words))

['a', "a's", 'able', 'about', 'above', 'according', 'accordingly', 'across', 'actual',
=====
taille de la liste stop_words: 572
```

```
In [78]: words_nostop = []
         for tweet in wordsnoslang2:
             for t in tweet:
                 if t.lower in stop_words:
                     tweet.remove(t)
             words_nostop.append(tweet)
         print(words_nostop)

[['stellargirl', 'I', 'love', 'my', 'Kindle2', 'Not', 'that', 'the', 'DX', 'is', 'o
```

was here', ['wrestling girl loves Khadda @ Not a Keds HAT Eldis déantat / scarwynprights 'Tim Bewatibg Kaggdée

```
In [84]: X_train, X_test, y_train, y_test = train_test_split(all_words, y_true, tra
```

```
In [90]: pipeline_pos = Pipeline([('vectorizer', CountVectorizer()), ('nb_sk', Mult

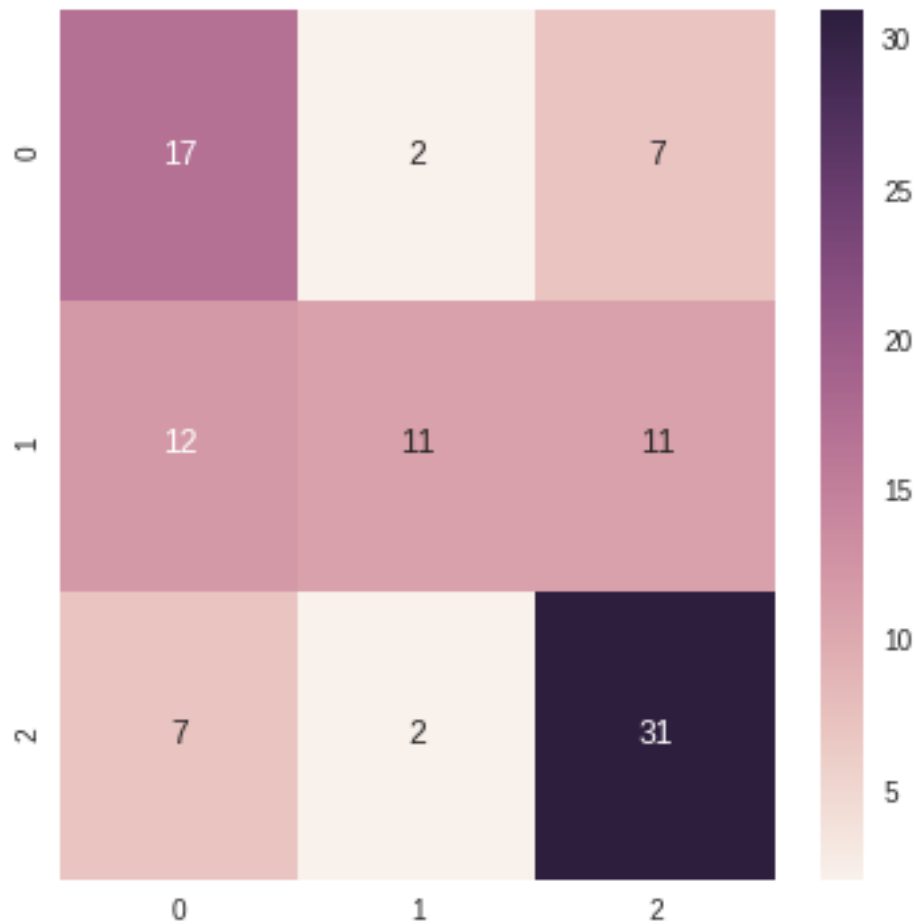
    scores_pos = cross_val_score(pipeline_pos, X_train, y_train, cv=5)
    pipeline_pos.fit(X_train, y_train)
    y_pred3 = pipeline_pos.predict(X_test)
    print(scores_pos)
    print(' ')
    print('la moyenne des scores obtenus par cross_validation sur le train est
    print('la déviation standard sur le train est:', scores_pos.std())
    print(' ')
    print(classification_report(y_test,y_pred3))
```

[ 0.61728395 0.675 0.73417722 0.67088608 0.6835443 ]

la moyenne des scores obtenus par cross\_validation sur le train est: 0.676178309111  
la déviation standard sur le train est: 0.0371915920354

	precision	recall	f1-score	support
-1	0.47	0.65	0.55	26
0	0.73	0.32	0.45	34
1	0.63	0.78	0.70	40
avg / total	0.63	0.59	0.57	100

```
In [92]: conf_matrix_meth03 = confusion_matrix(y_test, y_pred3)
    fig = plt.figure(figsize=(6,6))
    ax = sns.heatmap(conf_matrix_meth03, annot=True)
```



**Conclusion:** Ici, un simple Naïve Bayes est bien meilleur pour classer les tweets par rapport aux méthodes vues précédemment.

On aurait pu également utiliser TFIDF dans ce dernier cas.

Pour les premières méthodes avec Sentiword, il serait possible d'utiliser la distance de Levenshtein pour choisir correctement le bon score d'un mot (par exemple pour I, il est proposé en premier choix Iodine et non le pronom).