

preprocess_gensim

February 5, 2021

0.1 imports

```
[1]: #importing packages
import warnings
warnings.filterwarnings("ignore")

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from wordcloud import WordCloud
import seaborn as sns

# import re
import string

# nltk
import nltk
# from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords, wordnet
from nltk.stem.porter import *
from nltk.stem import WordNetLemmatizer

# sklearn
from sklearn.decomposition import PCA
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import LinearSVC, SVC
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, classification_report

# gensim
from gensim.models import Word2Vec

%matplotlib inline
```

train data: <https://www.kaggle.com/arkhoshghalb/twitter-sentiment-analysis-hatred-speech>

```
[2]:
```

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is s...
1	2	0	@user @user thanks for #lyft credit i can't us...
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in ...
4	5	0	factsguide: society now #motivation
...
31957	31958	0	ate @user isz that youuu?Ã°Ã°Ã°Ã°Ã°Ã°Ã°Ã°Ã°...
31958	31959	0	to see nina turner on the airwaves trying to...
31959	31960	0	listening to sad songs on a monday morning otw...
31960	31961	1	@user #sikh #temple vandalised in in #calgary,...
31961	31962	0	thank you @user for you follow

```
[4]: sentences = combi['tweet'].tolist()
      sentences_ss = " ".join(sentences)

      wordcloud = WordCloud(width=800, height=500, random_state=21,
                             max_font_size=110).generate(sentences_ss)
      plt.figure(figsize=(10, 7))
      plt.imshow(wordcloud, interpolation="bilinear")
      plt.axis('off')
      plt.show()
```



```
[6]:      id  label                                tweet \
0    1    0.0    @user when a father is dysfunctional and is s...
1    2    0.0    @user @user thanks for #lyft credit i can't us...
2    3    0.0                                bihday your majesty
3    4    0.0    #model    i love u take with u all the time in ...
4    5    0.0                                factsguide: society now    #motivation
```

```
                                cleanTweet
0      when a father is dysfunctional and is so se...
1      thanks for #lyft credit i can't use cause ...
2                                bihday your majesty
3    #model    i love u take with u all the time in ...
4                                factsguide: society now    #motivation
```

```
[7]: #upper case to lower case
combi['cleanTweet'] = combi['cleanTweet'].map(lambda x: x.lower())
#remove number
combi['cleanTweet'] = combi['cleanTweet'].map(lambda x: re.sub(r'\d+', '', x))
#remove punctuation
combi['cleanTweet'] = combi['cleanTweet'].map(lambda x: x.translate(x.
    ↳ maketrans('', '', string.punctuation)))
#remove whitespace
combi['cleanTweet'] = combi['cleanTweet'].map(lambda x: x.strip())
#remove url
url_cleaner = "https?:\S+|http?:\S|[\^A-Za-z0-9]+"
combi['cleanTweet'] = combi['cleanTweet'].map(lambda x: re.sub(url_cleaner, '␣
    ↳ ', x))
#removing small words
combi['cleanTweet'] = combi['cleanTweet'].apply(lambda x: ' '.join([w for w in␣
    ↳ x.split() if len(w)>3]))
# Replace 3 or more consecutive letters by 2 letter.
combi['cleanTweet'] = combi['cleanTweet'].map(lambda x: re.sub(r"(\.)\1\1+",␣
    ↳ r"\1\1", x))
combi
```

```
[7]:      id  label                                tweet \
0      1    0.0    @user when a father is dysfunctional and is s...
1      2    0.0    @user @user thanks for #lyft credit i can't us...
2      3    0.0                                bihday your majesty
3      4    0.0    #model    i love u take with u all the time in ...
4      5    0.0                                factsguide: society now    #motivation
...    ...    ...                                ...
49154  49155    NaN    thought factory: left-right polarisation! #tru...
49155  49156    NaN    feeling like a mermaid Ã°Ã°Ã°Ã° #hairflip #nev...
49156  49157    NaN    #hillary #campaigned today in #ohio((omg)) &am...
49157  49158    NaN    happy, at work conference: right mindset leads...
49158  49159    NaN    my    song "so glad" free download!    #shoegaze ...
```

```

                                cleanTweet
0      when father dysfunctional selfish drags kids i...
1      thanks lyft credit cant cause they dont offer ...
2                                bihday your majesty
3                                model love take with time
4                                factsguide society motivation
...
49154  thought factory leftright polarisation trump u...
49155  feeling like mermaid hairflip neverready forma...
49156  hillary campaigned today ohioomg used words li...
49157  happy work conference right mindset leads cult...
49158  song glad free download shoegaze newmusic newsong

[49159 rows x 4 columns]

```

```

[8]: #Tokenization
tokenized_tweet = combi['cleanTweet'].apply(lambda x: x.split())
tokenized_tweet.head()

```

```

[8]: 0      [when, father, dysfunctional, selfish, drags, ...
1      [thanks, lyft, credit, cant, cause, they, dont...
2                                [bihday, your, majesty]
3                                [model, love, take, with, time]
4                                [factsguide, society, motivation]
Name: cleanTweet, dtype: object

```

```

[9]: # Removing Consecutive letters
words_with_consecutive_letters = ['little', 'called', 'better', 'common',
→ 'cannot', 'matter', 'middle',
'follow', 'choose', 'letter', 'killed', 'summer', 'battle', 'supply',
→ 'bottom', 'passed', 'button',
'manner', 'happen', 'speech', 'troops', 'filled', 'latter', 'yellow',
→ 'dinner', 'coffee', 'fellow', 'tissue',
'hidden', 'smooth', 'pulled', 'narrow', 'lesson', 'missed', 'cattle',
→ 'winner', 'bigger', 'valley', 'dollar',
'suffer', 'mirror', 'passes', 'losses', 'terror', 'cotton', 'summit',
→ 'cheese', 'bottle', 'freely', 'vessel',
'sheets', 'sudden', 'copper', 'fallen', 'differ', 'gotten', 'buffer',
→ 'banner', 'ball',
'been', 'beer', 'beet', 'beep', 'bell', 'boom', 'boot', 'book', 'bull',
→ 'butt', 'call', 'cell', 'coon',
'dell', 'doll', 'door', 'doom', 'fall', 'fell', 'feel', 'feet', 'foot',
→ 'food', 'fool', 'fuss',
'full', 'gull', 'gall', 'hall', 'hell', 'heed', 'heel', 'hill', 'hull', 'all',
→ 'add', 'bee', 'boo',

```

```

'ell', 'ebb', 'egg', 'fee', 'goo', 'too', 'tee', 'see', 'Hollywood',
↳'bookkeeping', 'aloofness',
'foolishness', 'woodcutter', 'bulletproof', 'bootlegging', 'bittersweet',
↳'schoolfellow', 'committee',
'wheelbarrow', 'boorishness', 'bookkeeper', 'beerbelly', 'look', 'tomorrow',
↳'will', 'school', 'weekend']

tokenized_tweet = tokenized_tweet.apply(lambda text: [word if word in
↳words_with_consecutive_letters else re.sub(r"(\.|\1+)", r"\1", word) for word
↳in text])
tokenized_tweet.head()

```

```

[9]: 0    [when, father, dysfunctional, selfish, drags, ...
     1    [thanks, lyft, credit, cant, cause, they, dont...
     2                                [bihday, your, majesty]
     3                                [model, love, take, with, time]
     4                                [factsguide, society, motivation]
Name: cleanTweet, dtype: object

```

```

[10]: # Removing stop words
stop_words = set(stopwords.words('english'))
tokenized_tweet = tokenized_tweet.apply(lambda text: [word for word in text if
↳word not in stop_words])
tokenized_tweet.head()

```

```

[10]: 0    [father, dysfunctional, selfish, drags, kids, ...
     1    [thanks, lyft, credit, cant, cause, dont, ofer...
     2                                [bihday, majesty]
     3                                [model, love, take, time]
     4                                [factsguide, society, motivation]
Name: cleanTweet, dtype: object

```

```

[11]: # stemming
stemmer = PorterStemmer()
tokenized_tweet = tokenized_tweet.apply(lambda x: [stemmer.stem(i) for i in x])
tokenized_tweet.head()

```

```

[11]: 0    [father, dysfunct, selfish, drag, kid, dysfunct]
     1    [thank, lyft, credit, cant, caus, dont, ofer, ...
     2                                [bihday, majesti]
     3                                [model, love, take, time]
     4                                [factsguid, societi, motiv]
Name: cleanTweet, dtype: object

```

```

[12]: # Lemmatizing the words using WordNet
lemmatizer = WordNetLemmatizer()

```

```
wordnet_map = {"N":wordnet.NOUN, "V":wordnet.VERB, "J":wordnet.ADJ, "R":wordnet.
↳ADV}
def lemmatize_words(text):
    pos_tagged_text = nltk.pos_tag(text)
    # pos_tagged_text = text.apply(lambda x: nltk.pos_tag(x))
    return [lemmatizer.lemmatize(word, wordnet_map.get(pos[0], wordnet.NOUN))
↳for word, pos in pos_tagged_text]

tokenized_tweet = tokenized_tweet.apply(lambda text: lemmatize_words(text))
tokenized_tweet.head()
```

```
[12]: 0      [father, dysfunct, selfish, drag, kid, dysfunct]
      1      [thank, lyft, credit, cant, caus, dont, ofer, ...
      2              [bihday, majesti]
      3              [model, love, take, time]
      4              [factsguid, societi, motiv]
      Name: cleanTweet, dtype: object
```

```
[13]: for i in range(len(tokenized_tweet)):
      tokenized_tweet[i] = ' '.join(tokenized_tweet[i])
      combi['cleanTweet'] = tokenized_tweet
      combi
```

```
[13]:
```

	id	label	tweet \
0	1	0.0	@user when a father is dysfunctional and is s...
1	2	0.0	@user @user thanks for #lyft credit i can't us...
2	3	0.0	bihday your majesty
3	4	0.0	#model i love u take with u all the time in ...
4	5	0.0	factsguide: society now #motivation
...
49154	49155	NaN	thought factory: left-right polarisation! #tru...
49155	49156	NaN	feeling like a mermaid ã°ã°ã°ã° #hairflip #nev...
49156	49157	NaN	#hillary #campaigned today in #ohio((omg)) &am...
49157	49158	NaN	happy, at work conference: right mindset leads...
49158	49159	NaN	my song "so glad" free download! #shoegaze ...
			cleanTweet
0			father dysfunct selfish drag kid dysfunct
1			thank lyft credit cant caus dont ofer whelchai...
2			bihday majesti
3			model love take time
4			factsguid societi motiv
...			...
49154			think factori leftright polaris trump uselect ...
49155			fele like mermaid hairflip nevereadi formal we...
49156			hilari campaign today ohioimg use word like ase...
49157			hapi work confer right mindset lead cultureofd...

```
[49159 rows x 4 columns]
```

[illegible]

```
[16]: # Creating the model and setting values for the various parameters
num_features = 400 # Word vector dimensionality
min_word_count = 1 # Minimum word count
num_workers = 6 # Number of parallel threads
context = 10 # Context window size
downsampling = 1e-3 # (0.001) Downsample setting for frequent words
```



```

tokenized_tweet = combi['cleanTweet'].apply(lambda x: x.split())

# print("Training model....")
w2v_model = ↳ Word2Vec(sentences=tokenized_tweet, workers=num_workers, min_count=min_word_count, window=cont
# model.build_vocab(data['cleanTweet'])

```

```

2021-02-05 02:38:24,701 : INFO : collecting all words and their counts
2021-02-05 02:38:24,702 : INFO : PROGRESS: at sentence #0, processed 0 words,
keeping 0 word types
2021-02-05 02:38:24,721 : INFO : PROGRESS: at sentence #10000, processed 66722
words, keeping 14498 word types
2021-02-05 02:38:24,739 : INFO : PROGRESS: at sentence #20000, processed 132681
words, keeping 22779 word types
2021-02-05 02:38:24,753 : INFO : PROGRESS: at sentence #30000, processed 198922
words, keeping 29708 word types
2021-02-05 02:38:24,766 : INFO : PROGRESS: at sentence #40000, processed 264796
words, keeping 36033 word types
2021-02-05 02:38:24,781 : INFO : collected 41206 word types from a corpus of
325639 raw words and 49159 sentences
2021-02-05 02:38:24,781 : INFO : Loading a fresh vocabulary
2021-02-05 02:38:24,846 : INFO : effective_min_count=1 retains 41206 unique
words (100% of original 41206, drops 0)
2021-02-05 02:38:24,847 : INFO : effective_min_count=1 leaves 325639 word corpus
(100% of original 325639, drops 0)
2021-02-05 02:38:24,966 : INFO : deleting the raw counts dictionary of 41206
items
2021-02-05 02:38:24,967 : INFO : sample=0.001 downsamples 34 most-common words
2021-02-05 02:38:24,967 : INFO : downsampling leaves estimated 311758 word
corpus (95.7% of prior 325639)
2021-02-05 02:38:25,034 : INFO : estimated required memory for 41206 words and
100 dimensions: 53567800 bytes
2021-02-05 02:38:25,034 : INFO : resetting layer weights
2021-02-05 02:38:33,782 : INFO : training model with 6 workers on 41206
vocabulary and 100 features, using sg=0 hs=0 sample=0.001 negative=5 window=10
2021-02-05 02:38:34,033 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:34,038 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:34,042 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:34,045 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:34,047 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:34,049 : INFO : worker thread finished; awaiting finish of 0
more threads

```

2021-02-05 02:38:34,049 : INFO : EPOCH - 1 : training on 325639 raw words
(311668 effective words) took 0.3s, 1204019 effective words/s
2021-02-05 02:38:34,349 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:34,350 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:34,355 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:34,356 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:34,363 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:34,369 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:34,369 : INFO : EPOCH - 2 : training on 325639 raw words
(311605 effective words) took 0.3s, 994387 effective words/s
2021-02-05 02:38:34,587 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:34,593 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:34,597 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:34,598 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:34,600 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:34,604 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:34,604 : INFO : EPOCH - 3 : training on 325639 raw words
(311674 effective words) took 0.2s, 1366721 effective words/s
2021-02-05 02:38:34,877 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:34,878 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:34,878 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:34,893 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:34,898 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:34,901 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:34,901 : INFO : EPOCH - 4 : training on 325639 raw words
(311653 effective words) took 0.3s, 1073567 effective words/s
2021-02-05 02:38:35,199 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:35,203 : INFO : worker thread finished; awaiting finish of 4
more threads

```

2021-02-05 02:38:35,205 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:35,213 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:35,214 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:35,217 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:35,217 : INFO : EPOCH - 5 : training on 325639 raw words
(311857 effective words) took 0.3s, 1011172 effective words/s
2021-02-05 02:38:35,218 : INFO : training on a 1628195 raw words (1558457
effective words) took 1.4s, 1086255 effective words/s

```

```

[17]: w2v_model.train(combi['cleanTweet'], total_examples=w2v_model.corpus_count,
↳ epochs=60)

# To make the model memory efficient
w2v_model.init_sims(replace=False)

# Saving the model for later use. Can be loaded using Word2Vec.load()
model_name = "word2vec"
w2v_model.save(model_name)

```

```

2021-02-05 02:38:35,222 : WARNING : Effective 'alpha' higher than previous
training cycles
2021-02-05 02:38:35,222 : INFO : training model with 6 workers on 41206
vocabulary and 100 features, using sg=0 hs=0 sample=0.001 negative=5 window=10
2021-02-05 02:38:35,550 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:35,551 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:35,551 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:35,552 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:35,553 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:35,553 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:35,554 : INFO : EPOCH - 1 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1826946 effective words/s
2021-02-05 02:38:35,889 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:35,890 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:35,891 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:35,892 : INFO : worker thread finished; awaiting finish of 2

```

more threads
2021-02-05 02:38:35,894 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:35,894 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:35,894 : INFO : EPOCH - 2 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1771603 effective words/s
2021-02-05 02:38:36,234 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:36,235 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:36,237 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:36,238 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:36,239 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:36,239 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:36,240 : INFO : EPOCH - 3 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1755198 effective words/s
2021-02-05 02:38:36,631 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:36,632 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:36,633 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:36,633 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:36,635 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:36,636 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:36,636 : INFO : EPOCH - 4 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1522130 effective words/s
2021-02-05 02:38:36,976 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:36,978 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:36,979 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:36,980 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:36,982 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:36,982 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:36,982 : INFO : EPOCH - 5 : training on 2117250 raw words

(597005 effective words) took 0.3s, 1744647 effective words/s
2021-02-05 02:38:37,313 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:37,314 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:37,315 : INFO : worker thread finished; awaiting finish of 3 more threads
2021-02-05 02:38:37,317 : INFO : worker thread finished; awaiting finish of 2 more threads
2021-02-05 02:38:37,318 : INFO : worker thread finished; awaiting finish of 1 more threads
2021-02-05 02:38:37,319 : INFO : worker thread finished; awaiting finish of 0 more threads
2021-02-05 02:38:37,319 : INFO : EPOCH - 6 : training on 2117250 raw words (597005 effective words) took 0.3s, 1798743 effective words/s
2021-02-05 02:38:37,655 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:37,656 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:37,658 : INFO : worker thread finished; awaiting finish of 3 more threads
2021-02-05 02:38:37,659 : INFO : worker thread finished; awaiting finish of 2 more threads
2021-02-05 02:38:37,660 : INFO : worker thread finished; awaiting finish of 1 more threads
2021-02-05 02:38:37,660 : INFO : worker thread finished; awaiting finish of 0 more threads
2021-02-05 02:38:37,661 : INFO : EPOCH - 7 : training on 2117250 raw words (597005 effective words) took 0.3s, 1765704 effective words/s
2021-02-05 02:38:37,998 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:37,999 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:38,000 : INFO : worker thread finished; awaiting finish of 3 more threads
2021-02-05 02:38:38,001 : INFO : worker thread finished; awaiting finish of 2 more threads
2021-02-05 02:38:38,001 : INFO : worker thread finished; awaiting finish of 1 more threads
2021-02-05 02:38:38,002 : INFO : worker thread finished; awaiting finish of 0 more threads
2021-02-05 02:38:38,003 : INFO : EPOCH - 8 : training on 2117250 raw words (597005 effective words) took 0.3s, 1764716 effective words/s
2021-02-05 02:38:38,339 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:38,341 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:38,343 : INFO : worker thread finished; awaiting finish of 3

more threads
2021-02-05 02:38:38,344 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:38,344 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:38,344 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:38,345 : INFO : EPOCH - 9 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1768556 effective words/s
2021-02-05 02:38:38,677 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:38,680 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:38,681 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:38,682 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:38,682 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:38,683 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:38,683 : INFO : EPOCH - 10 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1782470 effective words/s
2021-02-05 02:38:39,037 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:39,039 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:39,040 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:39,040 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:39,041 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:39,042 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:39,042 : INFO : EPOCH - 11 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1682572 effective words/s
2021-02-05 02:38:39,374 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:39,375 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:39,376 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:39,378 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:39,379 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:39,380 : INFO : worker thread finished; awaiting finish of 0

more threads
2021-02-05 02:38:39,380 : INFO : EPOCH - 12 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1792865 effective words/s
2021-02-05 02:38:39,721 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:39,722 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:39,723 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:39,724 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:39,725 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:39,725 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:39,726 : INFO : EPOCH - 13 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1746823 effective words/s
2021-02-05 02:38:40,086 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:40,086 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:40,088 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:40,090 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:40,091 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:40,092 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:40,092 : INFO : EPOCH - 14 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1652006 effective words/s
2021-02-05 02:38:40,502 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:40,503 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:40,503 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:40,504 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:40,506 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:40,507 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:40,507 : INFO : EPOCH - 15 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1451410 effective words/s
2021-02-05 02:38:40,903 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:40,905 : INFO : worker thread finished; awaiting finish of 4

more threads
2021-02-05 02:38:40,906 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:40,907 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:40,909 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:40,909 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:40,910 : INFO : EPOCH - 16 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1506380 effective words/s
2021-02-05 02:38:41,269 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:41,271 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:41,272 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:41,272 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:41,274 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:41,274 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:41,275 : INFO : EPOCH - 17 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1651042 effective words/s
2021-02-05 02:38:41,629 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:41,630 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:41,630 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:41,631 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:41,631 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:41,633 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:41,633 : INFO : EPOCH - 18 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1688971 effective words/s
2021-02-05 02:38:42,007 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:42,008 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:42,008 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:42,009 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:42,009 : INFO : worker thread finished; awaiting finish of 1

more threads
2021-02-05 02:38:42,010 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:42,010 : INFO : EPOCH - 19 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1600098 effective words/s
2021-02-05 02:38:42,371 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:42,372 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:42,373 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:42,374 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:42,375 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:42,376 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:42,376 : INFO : EPOCH - 20 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1651155 effective words/s
2021-02-05 02:38:42,750 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:42,752 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:42,753 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:42,753 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:42,755 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:42,756 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:42,756 : INFO : EPOCH - 21 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1585593 effective words/s
2021-02-05 02:38:43,222 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:43,225 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:43,226 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:43,226 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:43,227 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:43,228 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:43,228 : INFO : EPOCH - 22 : training on 2117250 raw words
(597005 effective words) took 0.5s, 1278462 effective words/s
2021-02-05 02:38:43,640 : INFO : worker thread finished; awaiting finish of 5

more threads
2021-02-05 02:38:43,641 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:43,641 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:43,642 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:43,644 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:43,644 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:43,645 : INFO : EPOCH - 23 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1452006 effective words/s
2021-02-05 02:38:43,997 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:43,999 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:44,000 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:44,000 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:44,001 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:44,001 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:44,002 : INFO : EPOCH - 24 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1690445 effective words/s
2021-02-05 02:38:44,346 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:44,348 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:44,351 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:44,352 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:44,353 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:44,353 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:44,354 : INFO : EPOCH - 25 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1711948 effective words/s
2021-02-05 02:38:44,730 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:44,730 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:44,732 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:44,732 : INFO : worker thread finished; awaiting finish of 2

more threads
2021-02-05 02:38:44,734 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:44,735 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:44,735 : INFO : EPOCH - 26 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1585874 effective words/s
2021-02-05 02:38:45,082 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:45,084 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:45,084 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:45,086 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:45,086 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:45,087 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:45,087 : INFO : EPOCH - 27 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1713974 effective words/s
2021-02-05 02:38:45,457 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:45,458 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:45,459 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:45,459 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:45,460 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:45,462 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:45,462 : INFO : EPOCH - 28 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1609587 effective words/s
2021-02-05 02:38:45,866 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:45,868 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:45,868 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:45,869 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:45,870 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:45,871 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:45,871 : INFO : EPOCH - 29 : training on 2117250 raw words

(597005 effective words) took 0.4s, 1480365 effective words/s
2021-02-05 02:38:46,212 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:46,212 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:46,213 : INFO : worker thread finished; awaiting finish of 3 more threads
2021-02-05 02:38:46,215 : INFO : worker thread finished; awaiting finish of 2 more threads
2021-02-05 02:38:46,216 : INFO : worker thread finished; awaiting finish of 1 more threads
2021-02-05 02:38:46,216 : INFO : worker thread finished; awaiting finish of 0 more threads
2021-02-05 02:38:46,217 : INFO : EPOCH - 30 : training on 2117250 raw words (597005 effective words) took 0.3s, 1747053 effective words/s
2021-02-05 02:38:46,554 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:46,555 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:46,556 : INFO : worker thread finished; awaiting finish of 3 more threads
2021-02-05 02:38:46,557 : INFO : worker thread finished; awaiting finish of 2 more threads
2021-02-05 02:38:46,558 : INFO : worker thread finished; awaiting finish of 1 more threads
2021-02-05 02:38:46,559 : INFO : worker thread finished; awaiting finish of 0 more threads
2021-02-05 02:38:46,559 : INFO : EPOCH - 31 : training on 2117250 raw words (597005 effective words) took 0.3s, 1762104 effective words/s
2021-02-05 02:38:46,924 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:46,925 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:46,927 : INFO : worker thread finished; awaiting finish of 3 more threads
2021-02-05 02:38:46,928 : INFO : worker thread finished; awaiting finish of 2 more threads
2021-02-05 02:38:46,928 : INFO : worker thread finished; awaiting finish of 1 more threads
2021-02-05 02:38:46,929 : INFO : worker thread finished; awaiting finish of 0 more threads
2021-02-05 02:38:46,929 : INFO : EPOCH - 32 : training on 2117250 raw words (597005 effective words) took 0.4s, 1628335 effective words/s
2021-02-05 02:38:47,279 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:47,280 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:47,281 : INFO : worker thread finished; awaiting finish of 3

more threads
2021-02-05 02:38:47,281 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:47,283 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:47,283 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:47,284 : INFO : EPOCH - 33 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1704685 effective words/s
2021-02-05 02:38:47,639 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:47,641 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:47,642 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:47,644 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:47,645 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:47,646 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:47,646 : INFO : EPOCH - 34 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1667590 effective words/s
2021-02-05 02:38:48,076 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:48,077 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:48,079 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:48,080 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:48,080 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:48,083 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:48,083 : INFO : EPOCH - 35 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1378150 effective words/s
2021-02-05 02:38:48,438 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:48,439 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:48,440 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:48,441 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:48,441 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:48,443 : INFO : worker thread finished; awaiting finish of 0

more threads
2021-02-05 02:38:48,443 : INFO : EPOCH - 36 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1682526 effective words/s
2021-02-05 02:38:48,804 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:48,805 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:48,807 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:48,808 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:48,809 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:48,809 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:48,810 : INFO : EPOCH - 37 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1645602 effective words/s
2021-02-05 02:38:49,154 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:49,156 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:49,157 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:49,158 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:49,158 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:49,160 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:49,160 : INFO : EPOCH - 38 : training on 2117250 raw words
(597005 effective words) took 0.3s, 1720090 effective words/s
2021-02-05 02:38:49,523 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:49,524 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:49,526 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:49,526 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:49,528 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:49,530 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:49,530 : INFO : EPOCH - 39 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1628008 effective words/s
2021-02-05 02:38:49,899 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:49,900 : INFO : worker thread finished; awaiting finish of 4

more threads
2021-02-05 02:38:49,901 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:49,901 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:49,903 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:49,903 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:49,904 : INFO : EPOCH - 40 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1620613 effective words/s
2021-02-05 02:38:50,262 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:50,263 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:50,264 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:50,264 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:50,265 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:50,265 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:50,265 : INFO : EPOCH - 41 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1666610 effective words/s
2021-02-05 02:38:50,673 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:50,677 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:50,678 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:50,679 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:50,680 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:50,681 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:50,682 : INFO : EPOCH - 42 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1449716 effective words/s
2021-02-05 02:38:51,079 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:51,080 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:51,080 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:51,082 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:51,083 : INFO : worker thread finished; awaiting finish of 1

more threads
2021-02-05 02:38:51,084 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:51,084 : INFO : EPOCH - 43 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1498229 effective words/s
2021-02-05 02:38:51,449 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:51,450 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:51,451 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:51,452 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:51,453 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:51,454 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:51,455 : INFO : EPOCH - 44 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1636539 effective words/s
2021-02-05 02:38:51,803 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:51,804 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:51,805 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:51,807 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:51,807 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:51,808 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:51,809 : INFO : EPOCH - 45 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1703755 effective words/s
2021-02-05 02:38:52,203 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:52,204 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:52,206 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:52,207 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:52,207 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:52,209 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:52,209 : INFO : EPOCH - 46 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1505877 effective words/s
2021-02-05 02:38:52,568 : INFO : worker thread finished; awaiting finish of 5

more threads
2021-02-05 02:38:52,569 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:52,570 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:52,572 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:52,573 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:52,573 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:52,574 : INFO : EPOCH - 47 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1655183 effective words/s
2021-02-05 02:38:52,924 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:52,925 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:52,926 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:52,926 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:52,926 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:52,927 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:52,928 : INFO : EPOCH - 48 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1703346 effective words/s
2021-02-05 02:38:53,290 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:53,293 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:53,294 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:53,294 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:53,295 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:53,296 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:53,296 : INFO : EPOCH - 49 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1634864 effective words/s
2021-02-05 02:38:53,666 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:53,667 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:53,671 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:53,672 : INFO : worker thread finished; awaiting finish of 2

more threads
2021-02-05 02:38:53,673 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:53,674 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:53,674 : INFO : EPOCH - 50 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1596319 effective words/s
2021-02-05 02:38:54,038 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:54,039 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:54,040 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:54,041 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:54,041 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:54,042 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:54,042 : INFO : EPOCH - 51 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1638312 effective words/s
2021-02-05 02:38:54,426 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:54,428 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:54,428 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:54,429 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:54,430 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:54,431 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:54,431 : INFO : EPOCH - 52 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1547141 effective words/s
2021-02-05 02:38:54,794 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:54,797 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:54,799 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:54,800 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:54,800 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:54,801 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:54,801 : INFO : EPOCH - 53 : training on 2117250 raw words

(597005 effective words) took 0.4s, 1632006 effective words/s
2021-02-05 02:38:55,140 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:55,141 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:55,143 : INFO : worker thread finished; awaiting finish of 3 more threads
2021-02-05 02:38:55,144 : INFO : worker thread finished; awaiting finish of 2 more threads
2021-02-05 02:38:55,144 : INFO : worker thread finished; awaiting finish of 1 more threads
2021-02-05 02:38:55,145 : INFO : worker thread finished; awaiting finish of 0 more threads
2021-02-05 02:38:55,145 : INFO : EPOCH - 54 : training on 2117250 raw words (597005 effective words) took 0.3s, 1751692 effective words/s
2021-02-05 02:38:55,498 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:55,501 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:55,503 : INFO : worker thread finished; awaiting finish of 3 more threads
2021-02-05 02:38:55,506 : INFO : worker thread finished; awaiting finish of 2 more threads
2021-02-05 02:38:55,506 : INFO : worker thread finished; awaiting finish of 1 more threads
2021-02-05 02:38:55,507 : INFO : worker thread finished; awaiting finish of 0 more threads
2021-02-05 02:38:55,507 : INFO : EPOCH - 55 : training on 2117250 raw words (597005 effective words) took 0.4s, 1663913 effective words/s
2021-02-05 02:38:55,923 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:55,925 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:55,925 : INFO : worker thread finished; awaiting finish of 3 more threads
2021-02-05 02:38:55,926 : INFO : worker thread finished; awaiting finish of 2 more threads
2021-02-05 02:38:55,927 : INFO : worker thread finished; awaiting finish of 1 more threads
2021-02-05 02:38:55,927 : INFO : worker thread finished; awaiting finish of 0 more threads
2021-02-05 02:38:55,928 : INFO : EPOCH - 56 : training on 2117250 raw words (597005 effective words) took 0.4s, 1437511 effective words/s
2021-02-05 02:38:56,304 : INFO : worker thread finished; awaiting finish of 5 more threads
2021-02-05 02:38:56,306 : INFO : worker thread finished; awaiting finish of 4 more threads
2021-02-05 02:38:56,307 : INFO : worker thread finished; awaiting finish of 3

more threads
2021-02-05 02:38:56,307 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:56,308 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:56,309 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:56,310 : INFO : EPOCH - 57 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1580503 effective words/s
2021-02-05 02:38:56,708 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:56,710 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:56,711 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:56,712 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:56,713 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:56,714 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:56,714 : INFO : EPOCH - 58 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1495383 effective words/s
2021-02-05 02:38:57,078 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:57,080 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:57,081 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:57,081 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:57,084 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:57,084 : INFO : worker thread finished; awaiting finish of 0
more threads
2021-02-05 02:38:57,084 : INFO : EPOCH - 59 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1629697 effective words/s
2021-02-05 02:38:57,509 : INFO : worker thread finished; awaiting finish of 5
more threads
2021-02-05 02:38:57,511 : INFO : worker thread finished; awaiting finish of 4
more threads
2021-02-05 02:38:57,512 : INFO : worker thread finished; awaiting finish of 3
more threads
2021-02-05 02:38:57,513 : INFO : worker thread finished; awaiting finish of 2
more threads
2021-02-05 02:38:57,514 : INFO : worker thread finished; awaiting finish of 1
more threads
2021-02-05 02:38:57,515 : INFO : worker thread finished; awaiting finish of 0

```

more threads
2021-02-05 02:38:57,516 : INFO : EPOCH - 60 : training on 2117250 raw words
(597005 effective words) took 0.4s, 1414304 effective words/s
2021-02-05 02:38:57,516 : INFO : training on a 127035000 raw words (35820300
effective words) took 22.3s, 1606765 effective words/s
2021-02-05 02:38:57,516 : INFO : precomputing L2-norms of word weight vectors
2021-02-05 02:38:57,531 : INFO : saving Word2Vec object under word2vec,
separately None
2021-02-05 02:38:57,531 : INFO : not storing attribute vectors_norm
2021-02-05 02:38:57,532 : INFO : not storing attribute cum_table
2021-02-05 02:38:58,167 : INFO : saved word2vec

```

```

[18]: def display_pca_scatterplot(model, words=None, sample=0):
    if words == None:
        if sample > 0:
            words = np.random.choice(list(model.vocab.keys()), sample)
        else:
            words = [ word for word in model.vocab ]

    word_vectors = np.array([model[w] for w in words])

    twodim = PCA().fit_transform(word_vectors)[:,:2]

    plt.figure(figsize=(10,10))
    plt.scatter(twodim[:,0], twodim[:,1], edgecolors='k', c='r')
    for word, (x,y) in zip(words, twodim):
        plt.text(x+0.05, y+0.05, word)
    plt.savefig('./plot.png', dpi=800)

```

0.4 Test word2vec

```

[19]: w2v_model.wv.most_similar("father")

```

```

[19]: [('fathersday', 0.9910413026809692),
      ('dad', 0.989273726940155),
      ('dadi', 0.9832088947296143),
      ('papa', 0.9753644466400146),
      ('daygod', 0.9721598625183105),
      ('dadsday', 0.9712474346160889),
      ('thehardwork', 0.9687293171882629),
      ('fatherday', 0.9684503078460693),
      ('lopez', 0.9644782543182373),
      ('lopezchiro', 0.9629985094070435)]

```

```

[20]: w2v_model.wv.vocab

```

```
[20]: {'father': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8a90>,  
'dysfunct': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8a50>,  
'selfish': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8650>,  
'drag': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8950>,  
'kid': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8c90>,  
'thank': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8bd0>,  
'lyft': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8c10>,  
'credit': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8cd0>,  
'cant': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8810>,  
'caus': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8790>,  
'dont': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8d10>,  
'ofer': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8b10>,  
'whelchair': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8b90>,  
'van': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8d50>,  
'disapoint': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8d90>,  
'gethank': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8dd0>,  
'bihday': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8e10>,  
'majesti': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8e50>,  
'model': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8e90>,  
'love': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8ed0>,  
'take': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8f10>,  
'time': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8f50>,  
'factsguid': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8f90>,  
'societi': <gensim.models.keyedvectors.Vocab at 0x7f77fbef8fd0>,  
'motiv': <gensim.models.keyedvectors.Vocab at 0x7f77f800050>,  
'huge': <gensim.models.keyedvectors.Vocab at 0x7f77f800090>,  
'fare': <gensim.models.keyedvectors.Vocab at 0x7f77f8000d0>,  
'talk': <gensim.models.keyedvectors.Vocab at 0x7f77f800110>,  
'leav': <gensim.models.keyedvectors.Vocab at 0x7f77f800150>,  
'chao': <gensim.models.keyedvectors.Vocab at 0x7f77f800190>,  
'disput': <gensim.models.keyedvectors.Vocab at 0x7f77f8001d0>,  
'alshowandnogo': <gensim.models.keyedvectors.Vocab at 0x7f77f800210>,  
'camp': <gensim.models.keyedvectors.Vocab at 0x7f77f800250>,  
'tomorrow': <gensim.models.keyedvectors.Vocab at 0x7f77f800290>,  
'dani': <gensim.models.keyedvectors.Vocab at 0x7f77f8002d0>,  
'next': <gensim.models.keyedvectors.Vocab at 0x7f77f800310>,  
'school': <gensim.models.keyedvectors.Vocab at 0x7f77f800350>,  
'year': <gensim.models.keyedvectors.Vocab at 0x7f77f800390>,  
'exam': <gensim.models.keyedvectors.Vocab at 0x7f77f8003d0>,  
'think': <gensim.models.keyedvectors.Vocab at 0x7f77f800410>,  
'hate': <gensim.models.keyedvectors.Vocab at 0x7f77f800450>,  
'imagin': <gensim.models.keyedvectors.Vocab at 0x7f77f800490>,  
'actorslif': <gensim.models.keyedvectors.Vocab at 0x7f77f8004d0>,  
'revolutionschol': <gensim.models.keyedvectors.Vocab at 0x7f77f800510>,  
'girl': <gensim.models.keyedvectors.Vocab at 0x7f77f800550>,  
'land': <gensim.models.keyedvectors.Vocab at 0x7f77f800590>,  
'alin': <gensim.models.keyedvectors.Vocab at 0x7f77f8005d0>,
```

'cav': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80610>,
 'champion': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80650>,
 'cleveland': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80690>,
 'clevelandcavali': <gensim.models.keyedvectors.Vocab at 0x7f77fbe806d0>,
 'welcom': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80710>,
 'ireland': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80750>,
 'consum': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80790>,
 'price': <gensim.models.keyedvectors.Vocab at 0x7f77fbe807d0>,
 'index': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80810>,
 'climb': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80850>,
 'previou': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80890>,
 'blog': <gensim.models.keyedvectors.Vocab at 0x7f77fbe808d0>,
 'silver': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80910>,
 'gold': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80950>,
 'forex': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80990>,
 'orlando': <gensim.models.keyedvectors.Vocab at 0x7f77fbe809d0>,
 'standwithorlando': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80a10>,
 'pulseshot': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80a50>,
 'orlandoshot': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80a90>,
 'bigerproblem': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80ad0>,
 'heabreak': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80b10>,
 'valu': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80b50>,
 'dadi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80b90>,
 'today': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80bd0>,
 'day': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80c10>,
 'getingf': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80c50>,
 'cal': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80c90>,
 'michigan': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80cd0>,
 'middl': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80d10>,
 'build': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80d50>,
 'wal': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80d90>,
 'chant': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80dd0>,
 'tcot': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80e10>,
 'coment': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80e50>,
 'australia': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80e90>,
 'opkilingbay': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80ed0>,
 'seashepherd': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80f10>,
 'helpcovedolphin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80f50>,
 'thecov': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80f90>,
 'ouchjunior': <gensim.models.keyedvectors.Vocab at 0x7f77fbe80fd0>,
 'angri': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82050>,
 'junior': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82090>,
 'yugyoem': <gensim.models.keyedvectors.Vocab at 0x7f77fbe820d0>,
 'paner': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82110>,
 'posit': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82150>,
 'retwet': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82190>,
 'agr': <gensim.models.keyedvectors.Vocab at 0x7f77fbe821d0>,

'friday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82210>,
'smile': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82250>,
'around': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82290>,
'user': <gensim.models.keyedvectors.Vocab at 0x7f77fbe822d0>,
'coki': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82310>,
'make': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82350>,
'peopl': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82390>,
'know': <gensim.models.keyedvectors.Vocab at 0x7f77fbe823d0>,
'esenti': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82410>,
'oil': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82450>,
'chemic': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82490>,
'euro': <gensim.models.keyedvectors.Vocab at 0x7f77fbe824d0>,
'blame': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82510>,
'conced': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82550>,
'goal': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82590>,
'roney': <gensim.models.keyedvectors.Vocab at 0x7f77fbe825d0>,
'give': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82610>,
'away': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82650>,
'fre': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82690>,
'kick': <gensim.models.keyedvectors.Vocab at 0x7f77fbe826d0>,
'bale': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82710>,
'littl': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82750>,
'dude': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82790>,
'baday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe827d0>,
'coneofsham': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82810>,
'cat': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82850>,
'pise': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82890>,
'funi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe828d0>,
'laugh': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82910>,
'product': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82950>,
'hapi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82990>,
'wine': <gensim.models.keyedvectors.Vocab at 0x7f77fbe829d0>,
'tol': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82a10>,
'who': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82a50>,
'weekend': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82a90>,
'open': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82ad0>,
'drink': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82b10>,
'lumpi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82b50>,
'say': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82b90>,
'prove': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82bd0>,
'tgif': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82c10>,
'gamedev': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82c50>,
'indiedev': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82c90>,
'indiegamedev': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82cd0>,
'squad': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82d10>,
'beauti': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82d50>,
'sign': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82d90>,

'vendor': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82dd0>,
'upsideoflorida': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82e10>,
'shopalysa': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82e50>,
'medium': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82e90>,
'presconfer': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82ed0>,
'antalya': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82f10>,
'turkey': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82f50>,
'sunday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82f90>,
'throwback': <gensim.models.keyedvectors.Vocab at 0x7f77fbe82fd0>,
'great': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85050>,
'panel': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85090>,
'mediat': <gensim.models.keyedvectors.Vocab at 0x7f77fbe850d0>,
'public': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85110>,
'servic': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85150>,
'go': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85190>,
'nightclub': <gensim.models.keyedvectors.Vocab at 0x7f77fbe851d0>,
'god': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85210>,
'night': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85250>,
'man': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85290>,
'action': <gensim.models.keyedvectors.Vocab at 0x7f77fbe852d0>,
'mean': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85310>,
'lose': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85350>,
'famili': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85390>,
'forev': <gensim.models.keyedvectors.Vocab at 0x7f77fbe853d0>,
'riporlando': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85410>,
'never': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85450>,
'chanc': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85490>,
'vote': <gensim.models.keyedvectors.Vocab at 0x7f77fbe854d0>,
'presidenti': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85510>,
'candid': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85550>,
'excit': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85590>,
'cycl': <gensim.models.keyedvectors.Vocab at 0x7f77fbe855d0>,
'lok': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85610>,
'difer': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85650>,
'alohafriday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85690>,
'exist': <gensim.models.keyedvectors.Vocab at 0x7f77fbe856d0>,
'positivevib': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85710>,
'hawaian': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85750>,
'fellow': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85790>,
'nohern': <gensim.models.keyedvectors.Vocab at 0x7f77fbe857d0>,
'sadley': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85810>,
'pas': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85850>,
'tonight': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85890>,
'gawa': <gensim.models.keyedvectors.Vocab at 0x7f77fbe858d0>,
'sing': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85910>,
'chere': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85950>,
'fire': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85990>,

'hard': <gensim.models.keyedvectors.Vocab at 0x7f77fbe859d0>,
 'monday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85a10>,
 'cloudi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85a50>,
 'weather': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85a90>,
 'disabl': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85ad0>,
 'oxygen': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85b10>,
 'godnight': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85b50>,
 'badmonday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85b90>,
 'unbeliev': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85bd0>,
 'centuri': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85c10>,
 'ned': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85c50>,
 'someth': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85c90>,
 'like': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85cd0>,
 'neverump': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85d10>,
 'xenophobia': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85d50>,
 'taylorswift': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85d90>,
 'bull': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85dd0>,
 'domin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85e10>,
 'direct': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85e50>,
 'whatev': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85e90>,
 'want': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85ed0>,
 'morn': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85f10>,
 'travelingram': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85f50>,
 'dalat': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85f90>,
 'ripinkylif': <gensim.models.keyedvectors.Vocab at 0x7f77fbe85fd0>,
 'word': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88050>,
 'tel': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88090>,
 'photoshop': <gensim.models.keyedvectors.Vocab at 0x7f77fbe880d0>,
 'enoughisenough': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88110>,
 'dontphotoshopeveryth': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88150>,
 'wheresalthenaturalphoto': <gensim.models.keyedvectors.Vocab at
 0x7f77fbe88190>,
 'cedarpoint': <gensim.models.keyedvectors.Vocab at 0x7f77fbe881d0>,
 'wait': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88210>,
 'hour': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88250>,
 'valravnn': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88290>,
 'line': <gensim.models.keyedvectors.Vocab at 0x7f77fbe882d0>,
 'stope': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88310>,
 'work': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88350>,
 'close': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88390>,
 'sunshin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe883d0>,
 'finali': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88410>,
 'finish': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88450>,
 'book': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88490>,
 'youv': <gensim.models.keyedvectors.Vocab at 0x7f77fbe884d0>,
 'awhil': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88510>,
 'bokworm': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88550>,

'ontothextnextnovel': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88590>,
'knick': <gensim.models.keyedvectors.Vocab at 0x7f77fbe885d0>,
'easy': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88610>,
'playof': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88650>,
'rol': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88690>,
'life': <gensim.models.keyedvectors.Vocab at 0x7f77fbe886d0>,
'social': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88710>,
'network': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88750>,
'embrac': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88790>,
'share': <gensim.models.keyedvectors.Vocab at 0x7f77fbe887d0>,
'snake': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88810>,
'echeveria': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88850>,
'blomsflow': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88890>,
'grow': <gensim.models.keyedvectors.Vocab at 0x7f77fbe888d0>,
'garden': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88910>,
'iphonesia': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88950>,
'bli': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88990>,
'blom': <gensim.models.keyedvectors.Vocab at 0x7f77fbe889d0>,
'basilicabotanica': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88a10>,
'amaz': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88a50>,
'afirm': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88a90>,
'whenev': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88ad0>,
'goe': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88b10>,
'wrong': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88b50>,
'fele': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88b90>,
'blue': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88bd0>,
'ilustr': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88c10>,
'best': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88c50>,
'gete': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88c90>,
'readi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88cd0>,
'remov': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88d10>,
'victum': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88d50>,
'pulseclub': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88d90>,
'prayfororlando': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88dd0>,
'nose': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88e10>,
'petunia': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88e50>,
'concelebr': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88e90>,
'albanpilgrimag': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88ed0>,
'first': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88f10>,
'scumbageri': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88f50>,
'begin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88f90>,
'super': <gensim.models.keyedvectors.Vocab at 0x7f77fbe88fd0>,
'zpamdelacruz': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c050>,
'wede': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c090>,
'dolor': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c0d0>,
'capa': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c110>,
'tarlac': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c150>,

'scourg': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c190>,
'play': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c1d0>,
'baroqu': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c210>,
'piec': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c250>,
'piano': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c290>,
'beyond': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c2d0>,
'belief': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c310>,
'let': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c350>,
'fight': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c390>,
'peac': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c3d0>,
'rayo': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c410>,
'video': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c450>,
'world': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c490>,
'hotvideo': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c4d0>,
'ascot': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c510>,
'babe': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c550>,
'fashion': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c590>,
'monochrom': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c5d0>,
'style': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c610>,
'instahapyday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c650>,
'wekendi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c690>,
'selfi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c6d0>,
'yolo': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c710>,
'xoxo': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c750>,
'likelik': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c790>,
'confer': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c7d0>,
'right': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c810>,
'mindset': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c850>,
'lead': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c890>,
'cultureofdevelop': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c8d0>,
'organ': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c910>,
'christina': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c950>,
'grimi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c990>,
'last': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8c9d0>,
'perform': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ca10>,
'shot': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ca50>,
'christinarip': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ca90>,
'voic': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cad0>,
'christinagrimi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cb10>,
'danc': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cb50>,
'roar': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cb90>,
'preschol': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cbd0>,
'student': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cc10>,
'proud': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cc50>,
'reali': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cc90>,
'wife': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ccd0>,
'ador': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cd10>,

'mi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cd50>,
'poland': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cd90>,
'show': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cdd0>,
'surgeri': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ce10>,
'name': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ce50>,
'bridget': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ce90>,
'she': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ced0>,
'everyth': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cf10>,
'jealou': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cf50>,
'chatiado': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cf90>,
'celebr': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8cfd0>,
'everi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e050>,
'fatherli': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e090>,
'role': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e0d0>,
'sure': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e110>,
'white': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e150>,
'establish': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e190>,
'folx': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e1d0>,
'rune': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e210>,
'promot': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e250>,
'greatn': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e290>,
'journey': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e2d0>,
'travel': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e310>,
'yeah': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e350>,
'thejourneybegin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e390>,
'helo': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e3d0>,
'hotwet': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e410>,
'venusexchang': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e450>,
'brochur': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e490>,
'ariv': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e4d0>,
'awork': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e510>,
'solut': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e550>,
'much': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e590>,
'stuf': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e5d0>,
'hapen': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e610>,
'florida': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e650>,
'shote': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e690>,
'disneygatoratack': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e6d0>,
'ferari': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e710>,
'sake': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e750>,
'championship': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e790>,
'clearli': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e7d0>,
'turn': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e810>,
'point': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e850>,
'ferarimerc': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e890>,
'ace': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e8d0>,
'test': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e910>,

'sek': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e950>,
'probe': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e990>,
'udtapunjab': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8e9d0>,
'leak': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ea10>,
'finger': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ea50>,
'amarind': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ea90>,
'wrape': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ead0>,
'senseaboutmath': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8eb10>,
'call': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8eb50>,
'race': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8eb90>,
'ident': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ebd0>,
'might': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ec10>,
'regurgit': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ec50>,
'cale': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ec90>,
'sometim': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ecd0>,
'rais': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ed10>,
'brow': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ed50>,
'golfstrengthandcondit': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ed90>,
'strong': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8edd0>,
'felixfoisgolf': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ee10>,
'greathonour': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ee50>,
'carerconvo': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ee90>,
'design': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8eed0>,
'inov': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ef10>,
'learn': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ef50>,
'space': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8ef90>,
'includ': <gensim.models.keyedvectors.Vocab at 0x7f77fbe8efd0>,
'wateringhol': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91050>,
'cave': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91090>,
'mountaintop': <gensim.models.keyedvectors.Vocab at 0x7f77fbe910d0>,
'campfir': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91110>,
'altright': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91150>,
'use': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91190>,
'insecur': <gensim.models.keyedvectors.Vocab at 0x7f77fbe911d0>,
'lure': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91210>,
'whitesupremaci': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91250>,
'cari': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91290>,
'wouldnt': <gensim.models.keyedvectors.Vocab at 0x7f77fbe912d0>,
'help': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91310>,
'control': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91350>,
'wont': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91390>,
'stop': <gensim.models.keyedvectors.Vocab at 0x7f77fbe913d0>,
'black': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91410>,
'market': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91450>,
'terror': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91490>,
'wors': <gensim.models.keyedvectors.Vocab at 0x7f77fbe914d0>,
'power': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91510>,

'mind': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91550>,
'heal': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91590>,
'bodi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe915d0>,
'altwaystoh': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91610>,
'healthi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91650>,
'woho': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91690>,
'wek': <gensim.models.keyedvectors.Vocab at 0x7f77fbe916d0>,
'place': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91710>,
'member': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91750>,
'rehears': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91790>,
'music': <gensim.models.keyedvectors.Vocab at 0x7f77fbe917d0>,
'look': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91810>,
'anounc': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91850>,
'midwek': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91890>,
'newmus': <gensim.models.keyedvectors.Vocab at 0x7f77fbe918d0>,
'watchthispac': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91910>,
'guitar': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91950>,
'chanel': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91990>,
'fuss': <gensim.models.keyedvectors.Vocab at 0x7f77fbe919d0>,
'watch': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91a10>,
'episod': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91a50>,
'oflin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91a90>,
'nice': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91ad0>,
'long': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91b10>,
'snapchat': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91b50>,
'redhead': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91b90>,
'vermillionr': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91bd0>,
'thing': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91c10>,
'incred': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91c50>,
'receiv': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91c90>,
'accept': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91cd0>,
'letter': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91d10>,
'master': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91d50>,
'back': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91d90>,
'octob': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91dd0>,
'godtim': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91e10>,
'histori': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91e50>,
'daughter': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91e90>,
'ride': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91ed0>,
'bike': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91f10>,
'driveway': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91f50>,
'enjoy': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91f90>,
'sumeim': <gensim.models.keyedvectors.Vocab at 0x7f77fbe91fd0>,
'memori': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94050>,
'station': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94090>,
'do': <gensim.models.keyedvectors.Vocab at 0x7f77fbe940d0>,
'cours': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94110>,

'alway': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94150>,
'hope': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94190>,
'gona': <gensim.models.keyedvectors.Vocab at 0x7f77fbe941d0>,
'happen': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94210>,
'anytim': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94250>,
'son': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94290>,
'coupl': <gensim.models.keyedvectors.Vocab at 0x7f77fbe942d0>,
'nake': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94310>,
'japanes': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94350>,
'hump': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94390>,
'humper': <gensim.models.keyedvectors.Vocab at 0x7f77fbe943d0>,
'edwardsvil': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94410>,
'pensylvania': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94450>,
'personalis': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94490>,
'shop': <gensim.models.keyedvectors.Vocab at 0x7f77fbe944d0>,
'col': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94510>,
'home': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94550>,
'truli': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94590>,
'sick': <gensim.models.keyedvectors.Vocab at 0x7f77fbe945d0>,
'trump': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94610>,
'obama': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94650>,
'resign': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94690>,
'chang': <gensim.models.keyedvectors.Vocab at 0x7f77fbe946d0>,
'nashviletour': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94710>,
'nashvileforev': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94750>,
'eurusd': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94790>,
'clear': <gensim.models.keyedvectors.Vocab at 0x7f77fbe947d0>,
'barrier': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94810>,
'jump': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94850>,
'fresh': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94890>,
'high': <gensim.models.keyedvectors.Vocab at 0x7f77fbe948d0>,
'friend': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94910>,
'stil': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94950>,
'wrap': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94990>,
'head': <gensim.models.keyedvectors.Vocab at 0x7f77fbe949d0>,
'fact': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94a10>,
'destroy': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94a50>,
'bake': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94a90>,
'eyelid': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94ad0>,
'vigilfororlando': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94b10>,
'harp': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94b50>,
'clonakilti': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94b90>,
'ihavenofriend': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94bd0>,
'someonecomewithm': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94c10>,
'relax': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94c50>,
'icon': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94c90>,
'woman': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94cd0>,

'sundaymorn': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94d10>,
'marilynmonro': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94d50>,
'interest': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94d90>,
'linguist': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94dd0>,
'doesnt': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94e10>,
'adr': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94e50>,
'racism': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94e90>,
'raciolinguist': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94ed0>,
'bring': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94f10>,
'belov': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94f50>,
'recov': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94f90>,
'apl': <gensim.models.keyedvectors.Vocab at 0x7f77fbe94fd0>,
'marvel': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98050>,
'song': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98090>,
'musica': <gensim.models.keyedvectors.Vocab at 0x7f77fbe980d0>,
'wed': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98110>,
'ripchristina': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98150>,
'adel': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98190>,
'vine': <gensim.models.keyedvectors.Vocab at 0x7f77fbe981d0>,
'mock': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98210>,
'brexit': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98250>,
'spel': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98290>,
'referendum': <gensim.models.keyedvectors.Vocab at 0x7f77fbe982d0>,
'comerzbank': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98310>,
'health': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98350>,
'benefit': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98390>,
'cucumb': <gensim.models.keyedvectors.Vocab at 0x7f77fbe983d0>,
'pinoy': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98410>,
'folowm': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98450>,
'iger': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98490>,
'instagod': <gensim.models.keyedvectors.Vocab at 0x7f77fbe984d0>,
'toradora': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98510>,
'anim': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98550>,
'animedit': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98590>,
'breakup': <gensim.models.keyedvectors.Vocab at 0x7f77fbe985d0>,
'alon': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98610>,
'lone': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98650>,
'lap': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98690>,
'pol': <gensim.models.keyedvectors.Vocab at 0x7f77fbe986d0>,
'pick': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98710>,
'membership': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98750>,
'form': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98790>,
'cotd': <gensim.models.keyedvectors.Vocab at 0x7f77fbe987d0>,
'polar': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98810>,
'bear': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98850>,
'live': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98890>,
'cold': <gensim.models.keyedvectors.Vocab at 0x7f77fbe988d0>,

'loke': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98910>,
 'get': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98950>,
 'muslim': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98990>,
 'asasin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe989d0>,
 'snapi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98a10>,
 'fotbal': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98a50>,
 'fringesqcbagsum': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98a90>,
 'nochebuena': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98ad0>,
 'lasvega': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98b10>,
 'vega': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98b50>,
 'strip': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98b90>,
 'fleuret': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98bd0>,
 'instaboyinstamansefieseptumfriend': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98c10>,
 'smilesblackandwhit': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98c50>,
 'branch': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98c90>,
 'raini': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98cd0>,
 'write': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98d10>,
 'tear': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98d50>,
 'fli': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98d90>,
 'bird': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98dd0>,
 'haiku': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98e10>,
 'micropoetri': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98e50>,
 'buton': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98e90>,
 'mail': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98ed0>,
 'preti': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98f10>,
 'jewelrymak': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98f50>,
 'driver': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98f90>,
 'femal': <gensim.models.keyedvectors.Vocab at 0x7f77fbe98fd0>,
 'mose': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a050>,
 'river': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a090>,
 'weston': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a0d0>,
 'kill': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a110>,
 'crew': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a150>,
 'afterpa': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a190>,
 'japan': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a1d0>,
 'madeinjapan': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a210>,
 'cute': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a250>,
 'cawai': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a290>,
 'strawberi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a2d0>,
 'tomato': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a310>,
 'arent': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a350>,
 'protest': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a390>,
 'republican': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a3d0>,
 'wonthey': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a410>,
 'fuher': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a450>,
 'find': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a490>,

'spend': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a4d0>,
'updat': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a510>,
'analyt': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a550>,
'photoftoday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a590>,
'anyon': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a5d0>,
'date': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a610>,
'becam': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a650>,
'corupt': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a690>,
'emailg': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a6d0>,
'hilari': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a710>,
'shame': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a750>,
'disgrac': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a790>,
'ridicul': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a7d0>,
'stupid': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a810>,
'even': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a850>,
'negligencewhi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a890>,
'sachintendulkar': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a8d0>,
'instal': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a910>,
'thbihday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a950>,
'bestfriend': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a990>,
'impoant': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9a9d0>,
'mater': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9aa10>,
'pohu': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9aa50>,
'chri': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9aa90>,
'evan': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9aad0>,
'actor': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ab10>,
'human': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ab50>,
'chrisevan': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ab90>,
'hea': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9abd0>,
'prayer': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ac10>,
'murder': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ac50>,
'demo': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ac90>,
'album': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9acd0>,
'newalbum': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ad10>,
'indi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ad50>,
'echob': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ad90>,
'lion': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9add0>,
'webmaret': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ae10>,
'comun': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ae50>,
'manag': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ae90>,
'nzdusd': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9aed0>,
'target': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9af10>,
'bad': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9af50>,
'ever': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9af90>,
'bihdayweksuck': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9afd0>,
'birthday': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d050>,
'blese': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d090>,

'ladi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d0d0>,
'hapiest': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d110>,
'disneysmagickingdom': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d150>,
'disney': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d190>,
'magickingdom': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d1d0>,
'disneyland': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d210>,
'kinda': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d250>,
'among': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d290>,
'exclaim': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d2d0>,
'blain': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d310>,
'hair': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d350>,
'nite': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d390>,
'listen': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d3d0>,
'freedom': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d410>,
'efo': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d450>,
'present': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d490>,
'mere': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d4d0>,
'store': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d510>,
'past': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d550>,
'theodor': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d590>,
'rosevelt': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d5d0>,
'yal': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d610>,
'aint': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d650>,
'comit': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d690>,
'trust': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d6d0>,
'faith': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d710>,
'micheleobama': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d750>,
'gorila': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d790>,
'racist': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d7d0>,
'newarkfestiv': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d810>,
'internet': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d850>,
'broken': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d890>,
'netflix': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d8d0>,
'mochithepug': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d910>,
'puglif': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d950>,
'thathen': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d990>,
'whatbusin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9d9d0>,
'usual': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9da10>,
'dick': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9da50>,
'suckin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9da90>,
'tequila': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dad0>,
'lovin': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9db10>,
'slut': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9db50>,
'justalil': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9db90>,
'butstilontim': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dbd0>,
'smaler': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dc10>,
'hand': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dc50>,

'bari': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dc90>,
'probabl': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dcd0>,
'lie': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dd10>,
'game': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dd50>,
'suck': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dd90>,
'golf': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ddd0>,
'everyon': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9de10>,
'godmorn': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9de50>,
'chaleng': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9de90>,
'ap': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9ded0>,
'claim': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9df10>,
'punjabi': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9df50>,
'drugadict': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9df90>,
'dare': <gensim.models.keyedvectors.Vocab at 0x7f77fbe9dfd0>,
'tracerequest': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0050>,
'send': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0090>,
'depest': <gensim.models.keyedvectors.Vocab at 0x7f77fbea00d0>,
'condol': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0110>,
'zimbabw': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0150>,
'touch': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0190>,
'mato': <gensim.models.keyedvectors.Vocab at 0x7f77fbea01d0>,
'chapon': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0210>,
'saturday': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0250>,
'afternon': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0290>,
'meet': <gensim.models.keyedvectors.Vocab at 0x7f77fbea02d0>,
'so': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0310>,
'bolt': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0350>,
'bloke': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0390>,
'bing': <gensim.models.keyedvectors.Vocab at 0x7f77fbea03d0>,
'bong': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0410>,
'dawn': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0450>,
'month': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0490>,
'se': <gensim.models.keyedvectors.Vocab at 0x7f77fbea04d0>,
'vfest': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0510>,
'summer': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0550>,
'shane': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0590>,
'robe': <gensim.models.keyedvectors.Vocab at 0x7f77fbea05d0>,
'watson': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0610>,
'cricker': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0650>,
'milion': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0690>,
'jewishsupremacist': <gensim.models.keyedvectors.Vocab at 0x7f77fbea06d0>,
'hear': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0710>,
'move': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0750>,
'player': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0790>,
'whoknow': <gensim.models.keyedvectors.Vocab at 0x7f77fbea07d0>,
'radio': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0810>,
'coldplay': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0850>,

'come': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0890>,
'kepe': <gensim.models.keyedvectors.Vocab at 0x7f77fbea08d0>,
'terrorist': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0910>,
'constitut': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0950>,
'anoth': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0990>,
'excus': <gensim.models.keyedvectors.Vocab at 0x7f77fbea09d0>,
'apeas': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0a10>,
'measur': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0a50>,
'gloucesterugbi': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0a90>,
'togeth': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0ad0>,
'fine': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0b10>,
'dandi': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0b50>,
'exuber': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0b90>,
'depres': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0bd0>,
'hal': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0c10>,
'halfempti': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0c50>,
'mikeashley': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0c90>,
'sposdirectsham': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0cd0>,
'perhaps': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0d10>,
'exempl': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0d50>,
'protect': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0d90>,
'workersright': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0dd0>,
'within': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0e10>,
'europ': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0e50>,
'euref': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0e90>,
'remain': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0ed0>,
'stela': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0f10>,
'princ': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0f50>,
'birdstelabadprinc': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0f90>,
'post': <gensim.models.keyedvectors.Vocab at 0x7f77fbea0fd0>,
'check': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5050>,
'care': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5090>,
'alreadi': <gensim.models.keyedvectors.Vocab at 0x7f77fbea50d0>,
'block': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5110>,
'ask': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5150>,
'paicularli': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5190>,
'dificult': <gensim.models.keyedvectors.Vocab at 0x7f77fbea51d0>,
'question': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5210>,
'devast': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5250>,
'news': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5290>,
'victim': <gensim.models.keyedvectors.Vocab at 0x7f77fbea52d0>,
'hapin': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5310>,
'state': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5350>,
'manner': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5390>,
'margaret': <gensim.models.keyedvectors.Vocab at 0x7f77fbea53d0>,
'runbeck': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5410>,
'quot': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5450>,

'inspir': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5490>,
 'safe': <gensim.models.keyedvectors.Vocab at 0x7f77fbea54d0>,
 'way': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5510>,
 'acn': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5550>,
 'creat': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5590>,
 'beat': <gensim.models.keyedvectors.Vocab at 0x7f77fbea55d0>,
 'makingbeat': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5610>,
 'thursday': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5650>,
 'istandard': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5690>,
 'akaipro': <gensim.models.keyedvectors.Vocab at 0x7f77fbea56d0>,
 'number': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5710>,
 'weddingpay': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5750>,
 'weddingdr': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5790>,
 'forward': <gensim.models.keyedvectors.Vocab at 0x7f77fbea57d0>,
 'atend': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5810>,
 'cipd': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5850>,
 'workshop': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5890>,
 'onemoreday': <gensim.models.keyedvectors.Vocab at 0x7f77fbea58d0>,
 'reach': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5910>,
 'folow': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5950>,
 'twitch': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5990>,
 'hypo': <gensim.models.keyedvectors.Vocab at 0x7f77fbea59d0>,
 'stream': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5a10>,
 'cost': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5a50>,
 'verbal': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5a90>,
 'abus': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5ad0>,
 'adult': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5b10>,
 'ten': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5b50>,
 'countri': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5b90>,
 'bomb': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5bd0>,
 'stadium': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5c10>,
 'swalscha': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5c50>,
 'dejeun': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5c90>,
 'lunch': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5cd0>,
 'amoureux': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5d10>,
 'ilovethesecret': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5d50>,
 'lawofatract': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5d90>,
 'quiz': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5dd0>,
 'suce': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5e10>,
 'libtard': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5e50>,
 'liber': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5e90>,
 'polit': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5ed0>,
 'pole': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5f10>,
 'rozthediva': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5f50>,
 'blackgirlspol': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5f90>,
 'bodyandpol': <gensim.models.keyedvectors.Vocab at 0x7f77fbea5fd0>,
 'argentina': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7050>,

'atack': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7090>,
'empti': <gensim.models.keyedvectors.Vocab at 0x7f77fbea70d0>,
'citi': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7110>,
'side': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7150>,
'selfserv': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7190>,
'hypocrit': <gensim.models.keyedvectors.Vocab at 0x7f77fbea71d0>,
'kep': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7210>,
'bounc': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7250>,
'ball': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7290>,
'nimrod': <gensim.models.keyedvectors.Vocab at 0x7f77fbea72d0>,
'presid': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7310>,
'nohcarolina': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7350>,
'relect': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7390>,
'money': <gensim.models.keyedvectors.Vocab at 0x7f77fbea73d0>,
'ksleg': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7410>,
'cannot': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7450>,
'could': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7490>,
'easili': <gensim.models.keyedvectors.Vocab at 0x7f77fbea74d0>,
'anywher': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7510>,
'alremgroup': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7550>,
'alyalshahawi': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7590>,
'photographerdubai': <gensim.models.keyedvectors.Vocab at 0x7f77fbea75d0>,
'mydubai': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7610>,
'sem': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7650>,
'montreal': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7690>,
'wash': <gensim.models.keyedvectors.Vocab at 0x7f77fbea76d0>,
'spirit': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7710>,
'enteain': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7750>,
'brisban': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7790>,
'wimbledon': <gensim.models.keyedvectors.Vocab at 0x7f77fbea77d0>,
'entrepreneur': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7810>,
'speaker': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7850>,
'confid': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7890>,
'gala': <gensim.models.keyedvectors.Vocab at 0x7f77fbea78d0>,
'moment': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7910>,
'bipolar': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7950>,
'frikiegirl': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7990>,
'lucki': <gensim.models.keyedvectors.Vocab at 0x7f77fbea79d0>,
'chinatown': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7a10>,
'mal': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7a50>,
'binondo': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7a90>,
'pi': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7ad0>,
'isnt': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7b10>,
'complet': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7b50>,
'uter': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7b90>,
'moron': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7bd0>,
'trash': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7c10>,

'america': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7c50>,
'dad': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7c90>,
'magnetherapi': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7cd0>,
'sebo': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7d10>,
'smilegsmil': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7d50>,
'torio': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7d90>,
'briliant': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7dd0>,
'keter': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7e10>,
'custom': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7e50>,
'screne': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7e90>,
'potp': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7ed0>,
'shock': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7f10>,
'treatment': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7f50>,
'toraya': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7f90>,
'alamindah': <gensim.models.keyedvectors.Vocab at 0x7f77fbea7fd0>,
'carieunderwod': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab050>,
'ber': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab090>,
'bridpo': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab0d0>,
'food': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab110>,
'beer': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab150>,
'fest': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab190>,
'campsit': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab1d0>,
'drunk': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab210>,
'starstruck': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab250>,
'repost': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab290>,
'propey': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab2d0>,
'ocean': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab310>,
'sydney': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab350>,
'teach': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab390>,
'felt': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab3d0>,
'pakistan': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab410>,
'boy': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab450>,
'leave': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab490>,
'countdown': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab4d0>,
'nolefam': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab510>,
'milan': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab550>,
'stone': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab590>,
'rise': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab5d0>,
'club': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab610>,
'there': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab650>,
'reason': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab690>,
'stay': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab6d0>,
'nobodi': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab710>,
'el': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab750>,
'crush': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab790>,
'wakow': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab7d0>,
'buz': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab810>,

'hokup': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab850>,
'engag': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab890>,
'clasez': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab8d0>,
'risay': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab910>,
'risaypl': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab950>,
'legendari': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab990>,
'unit': <gensim.models.keyedvectors.Vocab at 0x7f77fbeab9d0>,
'rest': <gensim.models.keyedvectors.Vocab at 0x7f77fbeaba10>,
'proven': <gensim.models.keyedvectors.Vocab at 0x7f77fbeaba50>,
'firearm': <gensim.models.keyedvectors.Vocab at 0x7f77fbeaba90>,
'hold': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabad0>,
'door': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabb10>,
'that': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabb50>,
'deni': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabb90>,
'jefreycwiliam': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabbd0>,
'butru': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabc10>,
'babi': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabc50>,
'babygirl': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabc90>,
'niec': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabcd0>,
'oneteamlincoln': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabd10>,
'carowind': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabd50>,
'rockthepark': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabd90>,
'timetogetmyjesuson': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabdd0>,
'buy': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabe10>,
'dori': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabe50>,
'ticket': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabe90>,
'minut': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabed0>,
'sixti': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabf10>,
'second': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabf50>,
'sanjapan': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabf90>,
'animeconvent': <gensim.models.keyedvectors.Vocab at 0x7f77fbeabfd0>,
'guy': <gensim.models.keyedvectors.Vocab at 0x7f77fbae050>,
'septemb': <gensim.models.keyedvectors.Vocab at 0x7f77fbae090>,
'moto': <gensim.models.keyedvectors.Vocab at 0x7f77fbae0d0>,
'tech': <gensim.models.keyedvectors.Vocab at 0x7f77fbae110>,
'count': <gensim.models.keyedvectors.Vocab at 0x7f77fbae150>,
'mondaymotiv': <gensim.models.keyedvectors.Vocab at 0x7f77fbae190>,
'livelifetothe fullest': <gensim.models.keyedvectors.Vocab at 0x7f77fbae1d0>,
'everydaycount': <gensim.models.keyedvectors.Vocab at 0x7f77fbae210>,
'degr': <gensim.models.keyedvectors.Vocab at 0x7f77fbae250>,
'ana': <gensim.models.keyedvectors.Vocab at 0x7f77fbae290>,
'clasmat': <gensim.models.keyedvectors.Vocab at 0x7f77fbae2d0>,
'would': <gensim.models.keyedvectors.Vocab at 0x7f77fbae310>,
'abl': <gensim.models.keyedvectors.Vocab at 0x7f77fbae350>,
'supo': <gensim.models.keyedvectors.Vocab at 0x7f77fbae390>,
'event': <gensim.models.keyedvectors.Vocab at 0x7f77fbae3d0>,
'brentwod': <gensim.models.keyedvectors.Vocab at 0x7f77fbae410>,

```

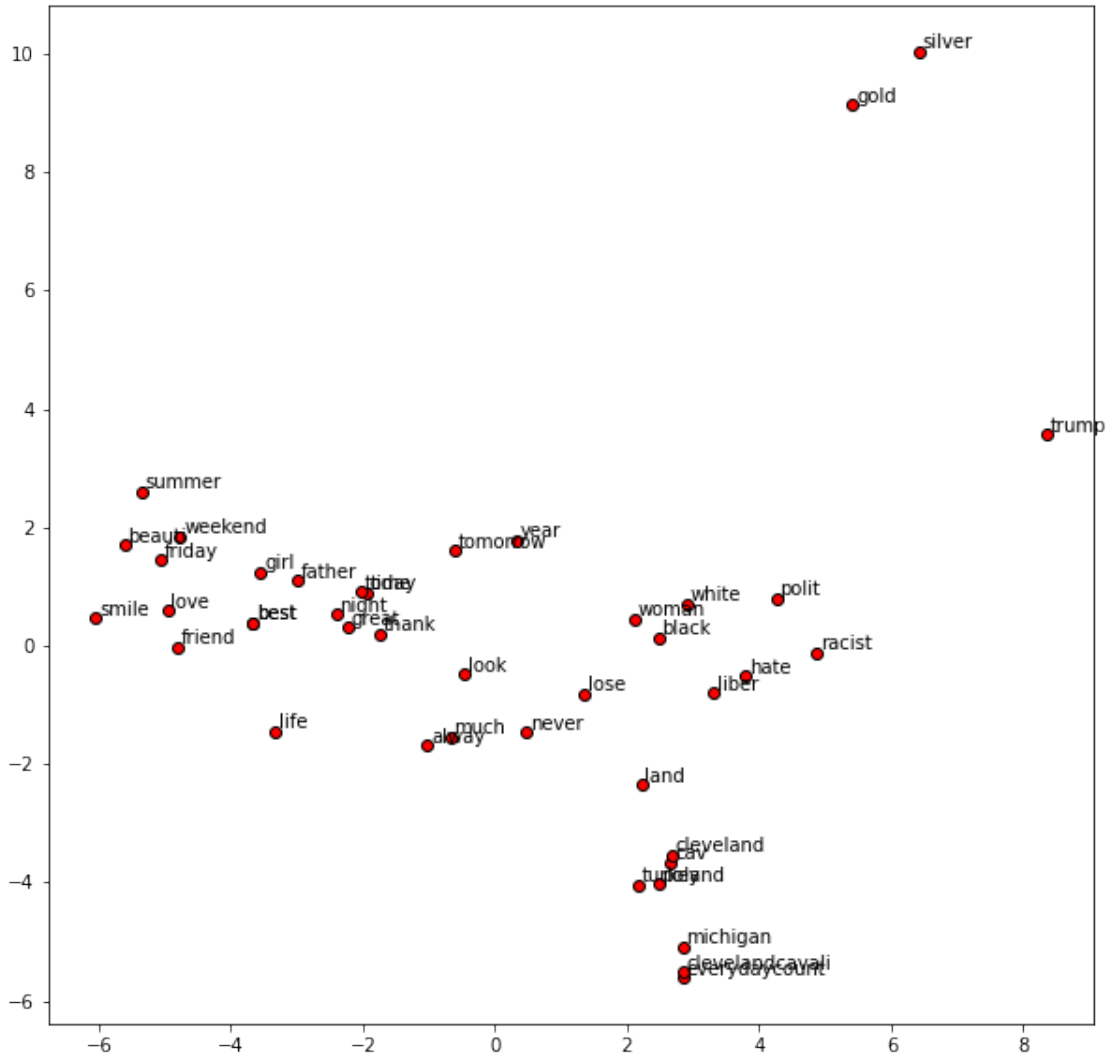
'homesta': <gensim.models.keyedvectors.Vocab at 0x7f77fbae450>,
'fund': <gensim.models.keyedvectors.Vocab at 0x7f77fbae490>,
'hapyhapi': <gensim.models.keyedvectors.Vocab at 0x7f77fbae4d0>,
'kiten': <gensim.models.keyedvectors.Vocab at 0x7f77fbae510>,
'blackiten': <gensim.models.keyedvectors.Vocab at 0x7f77fbae550>,
'opoun': <gensim.models.keyedvectors.Vocab at 0x7f77fbae590>,
'cherish': <gensim.models.keyedvectors.Vocab at 0x7f77fbae5d0>,
'themeowod': <gensim.models.keyedvectors.Vocab at 0x7f77fbae610>,
'pupi': <gensim.models.keyedvectors.Vocab at 0x7f77fbae650>,
'til': <gensim.models.keyedvectors.Vocab at 0x7f77fbae690>,
'pharel': <gensim.models.keyedvectors.Vocab at 0x7f77fbae6d0>,
'wiliam': <gensim.models.keyedvectors.Vocab at 0x7f77fbae710>,
'pharelwiliam': <gensim.models.keyedvectors.Vocab at 0x7f77fbae750>,
'lyric': <gensim.models.keyedvectors.Vocab at 0x7f77fbae790>,
'yesterday': <gensim.models.keyedvectors.Vocab at 0x7f77fbae7d0>,
...}

```

```

[21]: display_pca_scatterplot(w2v_model,
                                ['friend', 'great', 'love', 'life', 'girl', 'beauti',
                                ↪ 'smile',
                                'summer', 'thank', 'time', 'today', 'weekend',
                                ↪ 'father', 'polit',
                                'alway', 'year', 'best', 'tomorrow', 'land', 'cav',
                                ↪ 'white', 'woman',
                                'friday', 'night', 'never', 'trump', 'much', 'lose',
                                ↪ 'best', 'black',
                                'cleveland', 'poland', 'turkey', 'silver', 'gold',
                                ↪ 'everydaycount',
                                'hate', 'clevelandcavali', 'michigan', 'liber',
                                ↪ 'look', 'racist'])

```



0.5 TF-IDF

```
[22]: # tfidf_vectorizer = TfidfVectorizer(max_df=0.90, min_df=2, max_features=1000,
      ↪ stop_words='english')
tfidf_vectorizer = TfidfVectorizer()
# TF-IDF feature matrix
tfidf_matrix = tfidf_vectorizer.fit_transform(combi['cleanTweet'])
tfidf_feature_names = tfidf_vectorizer.get_feature_names()
```

```
[23]: w2v_model.wv.syn0.shape
```

```
[23]: (41206, 100)
```

```
[24]: def make_feature_vec(words, model, num_features, tfidf_scores):
    """
    Average the word vectors for a set of words
    """
    feature_vec = np.zeros((num_features,), dtype="float32") # pre-initialize
    ↪ (for speed)
    nwords = 0.
    index2word_set = set(model.index2word) # words known to the model

    for word in words:
        if word in index2word_set:
            nwords = nwords + 1.
            weight = 1
            if word in tfidf_scores:
                weight = tfidf_scores[word]
            feature_vec = np.add(feature_vec, model[word] * weight)

    feature_vec = np.divide(feature_vec, nwords)
    return feature_vec

def get_avg_feature_vecs(sents, model, num_features, tfidf_list):
    """
    Calculate average feature vectors for all reviews
    """
    counter = 0
    sent_feature_vecs = np.zeros((len(sents), num_features), dtype='float32') #
    ↪ pre-initialize (for speed)

    feature_index = tfidf_matrix[counter,:].nonzero()[1]

    for sent in sents:
        sent_feature_vecs[counter] = make_feature_vec(sent, model,
    ↪ num_features, tfidf_list[counter])
        counter = counter + 1
    return sent_feature_vecs

[25]: index_value={i[1]:i[0] for i in tfidf_vectorizer.vocabulary_.items()}
fully_indexed = []
for row in tfidf_matrix:
    fully_indexed.append({index_value[column]:value for (column,value) in
    ↪ zip(row.indices,row.data)})

data = get_avg_feature_vecs(tokenized_tweet, w2v_model.wv, 100, fully_indexed)
data
```

```
[25]: array([[ -0.02657441,  0.12916605,  0.09264615, ...,  0.01442369,
          0.08974385, -0.01524654],
          [-0.06197399,  0.1441003 ,  0.08470177, ..., -0.04684628,
          0.08766985,  0.02546517],
          [-0.14107616,  0.16237888,  0.16383454, ...,  0.02730435,
          0.06428895, -0.13175893],
          ...,
          [-0.07688612,  0.1844849 ,  0.09902883, ..., -0.04070211,
          0.08921685,  0.03672873],
          [-0.49621707,  0.43876347,  0.32817608, ...,  0.00303068,
          -0.03060175,  0.05374011],
          [-0.05516927,  0.11511093,  0.11614979, ..., -0.06228389,
          0.0706454 ,  0.03964414]], dtype=float32)
```

1 Classification Models

```
[26]: train_data = data[:31962]
      # test_data = data[25569:]
      # train_label = combi['label'].to_numpy()[:25569]
      # test_label = combi['label'].to_numpy()[25569:]

      # train_data.shape
      # np.isnan(train_data).any()
      train_data = np.nan_to_num(train_data)

      # splitting data into training and validation set
      X_train, X_test, y_train, y_test = train_test_split(train_data, train['label'],
      ↪random_state=42, test_size=0.3)
```

```
[27]: # def model_Evaluate(model, xtest):
      def modelResult(y_pred):

          # Predict values for Test dataset
          # y_pred = model.predict(xtest)

          # Print the evaluation metrics for the dataset.
          print(classification_report(y_test, y_pred))

          # Compute and plot the Confusion matrix
          cf_matrix = confusion_matrix(y_test, y_pred)

          categories = ['Negative', 'Positive']
          group_names = ['True Neg', 'False Pos', 'False Neg', 'True Pos']
          group_percentages = ['{0:.2%}'.format(value) for value in cf_matrix.
          ↪flatten() / np.sum(cf_matrix)]
```

```

labels = [f'{v1}\n{v2}' for v1, v2 in zip(group_names,group_percentages)]
labels = np.asarray(labels).reshape(2,2)

sns.heatmap(cf_matrix, annot = labels, cmap = 'Blues',fmt = '',
            xticklabels = categories, yticklabels = categories)

plt.xlabel("Predicted values", fontdict = {'size':14}, labelpad = 10)
plt.ylabel("Actual values", fontdict = {'size':14}, labelpad = 10)
plt.title ("Confusion Matrix", fontdict = {'size':18}, pad = 20)
plt.savefig("./result.png")

```

1.1 Random Forest

```

[28]: forest = RandomForestClassifier(n_estimators = 1000, random_state = 50)
forest = forest.fit(X_train, y_train)
forest_y_pred = forest.predict(X_test)

forest.score(X_test, y_test)

```

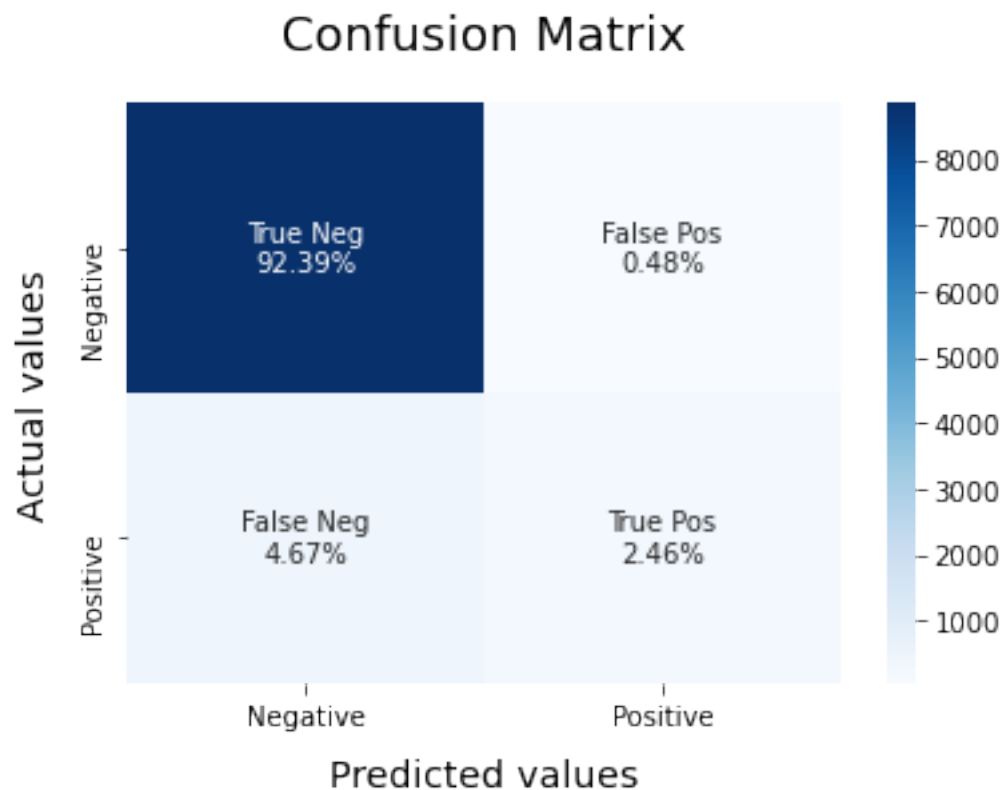
[28]: 0.9484826363541558

```

[29]: modelResult(forest_y_pred)

```

	precision	recall	f1-score	support
0	0.95	0.99	0.97	8905
1	0.84	0.35	0.49	684
accuracy			0.95	9589
macro avg	0.89	0.67	0.73	9589
weighted avg	0.94	0.95	0.94	9589



1.2 LinearSVC

```
[30]: pca = PCA(n_components=15)

lsvm_X_train = pca.fit_transform(X_train)
lsvm_X_test  = pca.transform(X_test)

SVCmodel = LinearSVC()
SVCmodel.fit(lsvm_X_train, y_train)
lsvm_y_pred = SVCmodel.predict(lsvm_X_test)

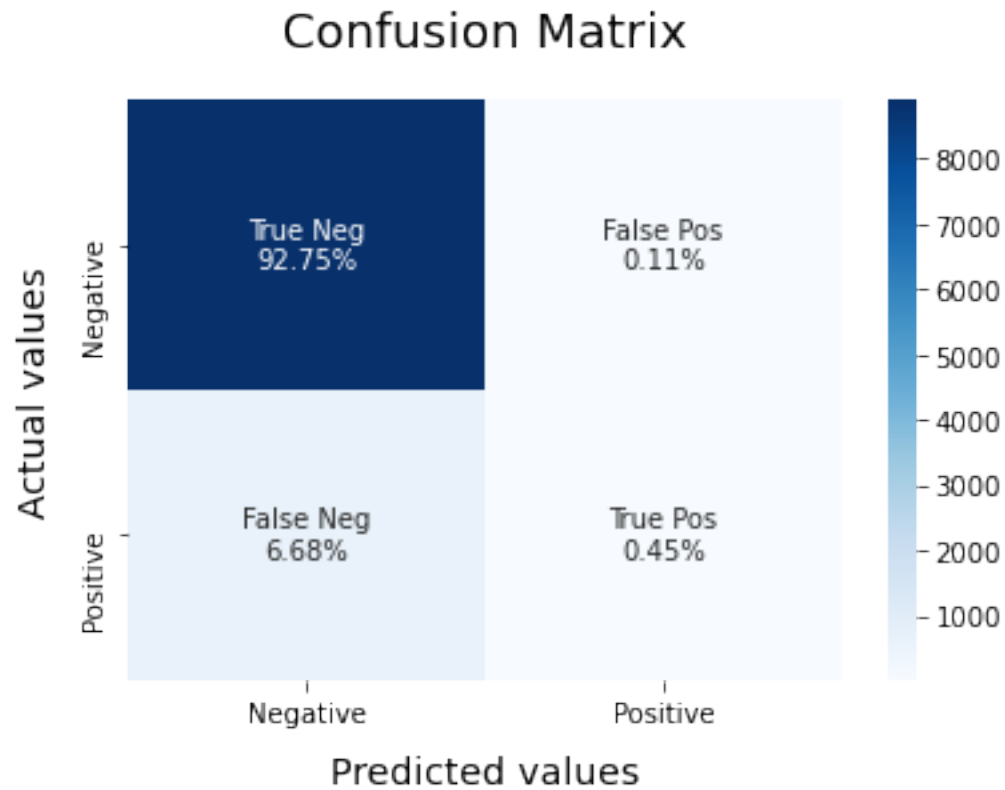
SVCmodel.score(lsvm_X_test, y_test)
```

[30]: 0.9320054228803838

```
[31]: modelResult(lsvm_y_pred)
```

	precision	recall	f1-score	support
0	0.93	1.00	0.96	8905
1	0.80	0.06	0.12	684

accuracy			0.93	9589
macro avg	0.86	0.53	0.54	9589
weighted avg	0.92	0.93	0.90	9589



1.3 svm with rbf

```
[32]: rbfSVCmodel = SVC(kernel='rbf', gamma=0.5, C=0.1)
      rbfSVCmodel.fit(X_train, y_train)
      rbf_svm_y_pred = rbfSVCmodel.predict(X_test)

      rbfSVCmodel.score(X_test, y_test)
```

[32]: 0.9311711335905726

```
[33]: modelResult(rbf_svm_y_pred)
```

	precision	recall	f1-score	support
0	0.93	1.00	0.96	8905

	1	1.00	0.04	0.07	684
accuracy				0.93	9589
macro avg		0.97	0.52	0.52	9589
weighted avg		0.94	0.93	0.90	9589

