**INVALID:(Python File)**

Importing the packages like nltk (natural language toolkit) and gensim

for removing stopwords using a corpus dictionary.

import warnings

warnings.filterwarnings(action='ignore', category=UserWarning, module='gensim')

import nltk

import gensim

from nltk.stem.wordnet import WordNetLemmatizer

from textblob import TextBlob

from nltk.corpus import stopwords

from nltk.stem.wordnet import WordNetLemmatizer

import string

stop = set(stopwords.words('english'))

invalid = set(stopwords.words('en'))

lemma = WordNetLemmatizer()

def get\_sentiment(tweet):

analysis = TextBlob(tweet)

filepath = 'E:/project/preprocess4/pp1/pp2/tq25.txt'

## Open the file with read only permit

f = open(filepath, encoding="utf8")

## use readlines to read all lines in the file

## The variable "lines" is a list containing all lines

lines = f.readlines()

sentiments=[None]\*20

for j in range(0,20):

s2=lines[j]

normalized = " ".join(lemma.lemmatize(word) for word in s2.split())

stop\_free = " ".join([i for i in normalized.lower().split() if i not in stop])

invalid\_free = " ".join([i for i in stop\_free.lower().split() if i in invalid])

lines[j]=invalid\_free + '\n'

sentiments[j]=get\_sentiment(lines[j])

# print(str(j)+"-- "+sentiments[j]+" -- "+lines[j])

outfilepath = 'E:/project/preprocess4/pp1/pp2/pp3/tr25.txt'

outf = open(outfilepath, "w")

outf.writelines(lines)

#sentimentfilepath = 'E:/project/preprocess1/senti\_m3.txt'

#sentimentf = open(sentimentfilepath, "w")

#for j in range(0,110):

# sentimentf.write(str(sentiments[j]+'\n'))

f.close()

outf.close()

#sentimentf.close()