Using the hashtag you received last week (NOT the person's name), do the following:

1. Connect the Twitter API with Python on your local computer.
2. Search the Twitter stream for the hashtag you have received
3. Collect 200 tweets in a data file.
4. Separate the datafile into a training set (50 tweets) and into a test set (150 tweets)
5. Go through the training set, review each tweet, and label it as either 'like' or 'dislike'
6. Then build a simple sentiment analysis (using Naive Bayes or the classification model shown in the bayesSentiment.py file in Chapter 7 Zacharski's Guide to Data Mining) that classifies the tweets in the test set as 'like' or 'dislike'

Twitter Api Connection and logithermic connection :  
  
  
Input:   
  
Input: 

from \_\_future\_\_ import print\_function

import os, codecs, math

class BayesText:

def \_\_init\_\_(self, trainingdir, stopwordlist, ignoreBucket):

"""This class implements a naive Bayes approach to text

classification

trainingdir is the training data. Each subdirectory of

trainingdir is titled with the name of the classification

category -- those subdirectories in turn contain the text

files for that category.

The stopwordlist is a list of words (one per line) will be

removed before any counting takes place.

"""

self.vocabulary = {}

self.prob = {}

self.totals = {}

self.stopwords = {}

f = open(stopwordlist)

for line in f:

self.stopwords[line.strip()] = 1

f.close()

categories = os.listdir(trainingdir)

#filter out files that are not directories

self.categories = [filename for filename in categories

if os.path.isdir(trainingdir + filename)]

print("\*\*\* init Function: Line 28 Categories and Counting ...\n\n")

for category in self.categories:

print(' ' + category) ## SHS: Verbose

(self.prob[category],

self.totals[category]) = self.train(trainingdir, category,

ignoreBucket)

# I am going to eliminate any word in the vocabulary

# that doesn't occur at least 3 times

toDelete = []

for word in self.vocabulary:

if self.vocabulary[word] < 3:

# mark word for deletion

# can't delete now because you can't delete

# from a list you are currently iterating over

toDelete.append(word)

# now delete

for word in toDelete:

del self.vocabulary[word]

# now compute probabilities

vocabLength = len(self.vocabulary)

print("\*\*\* init Function: Line 48 Computing probabilities:\n\n")

for category in self.categories:

print(' ' + category)

denominator = self.totals[category] + vocabLength

for word in self.vocabulary:

if word in self.prob[category]:

count = self.prob[category][word]

else:

count = 1

self.prob[category][word] = (float(count + 1)

/ denominator)

print ("\*\*\* Line 59 DONE TRAINING\n\n") ## SHS: uncommented verbose

def train(self, trainingdir, category, bucketNumberToIgnore):

"""counts word occurrences for a particular category"""

ignore = "%i" % bucketNumberToIgnore

currentdir = trainingdir + category

directories = os.listdir(currentdir)

print(directories) ## SHS checking variable

counts = {}

total = 0

print("\*\*\* Line 69 START TRAINING FUNCTION ...\n\n")

for directory in directories:

## print(currentdir + '/' + directory) ## SHS: added verbose

if directory != ignore:

## currentBucket = trainingdir + category + "/" + directory

currentBucket = currentdir + "/" + directory ## SHS: Cleaning up variables

files = os.listdir(currentBucket) ## throws error if hidden desktop.ini file is present

## print(" " + currentBucket) ## SHS: verbose

for file in files:

f = codecs.open(currentBucket + '/' + file, 'r', 'iso8859-1')

for line in f:

tokens = line.split()

for token in tokens:

# get rid of punctuation and lowercase token

token = token.strip('\'".,?:-')

token = token.lower()

if token != '' and not token in self.stopwords:

self.vocabulary.setdefault(token, 0)

self.vocabulary[token] += 1

counts.setdefault(token, 0)

counts[token] += 1

total += 1

f.close()

return(counts, total)

def classify(self, filename):

results = {}

for category in self.categories:

results[category] = 0

f = codecs.open(filename, 'r', 'iso8859-1')

print("\*\*\* Line 101: START CLASSIFICATION ...\n\n")

for line in f:

tokens = line.split()

for token in tokens:

print(token)

token = token.strip('\'".,?:-').lower()

if token in self.vocabulary:

for category in self.categories:

if self.prob[category][token] == 0:

print("%s %s" % (category, token))

results[category] += math.log(

self.prob[category][token])

f.close()

results = list(results.items())

results.sort(key=lambda tuple: tuple[1], reverse = True)

# for debugging I can change this to give me the entire list

return results[0][0]

def testCategory(self, direc, category, bucketNumber):

results = {}

directory = direc + ("%i/" % bucketNumber)

print("\*\*\* Line 122 START TESTING the categories now ...\n\n")

print("Testing " + directory)

files = os.listdir(directory)

total = 0

correct = 0

for file in files:

total += 1

result = self.classify(directory + file)

results.setdefault(result, 0)

results[result] += 1

#if result == category:

# correct += 1

return results

def test(self, testdir, bucketNumber):

"""Test all files in the test directory--that directory is

organized into subdirectories--each subdir is a classification

category"""

results = {}

categories = os.listdir(testdir)

#filter out files that are not directories

categories = [filename for filename in categories if

os.path.isdir(testdir + filename)]

correct = 0

total = 0

for category in categories:

#print(".", end="")

results[category] = self.testCategory(

testdir + category + '/', category, bucketNumber)

return results

def tenfold(dataPrefix, stoplist):

results = {}

for i in range(0,10):

bT = BayesText(dataPrefix, stoplist, i)

r = bT.test(theDir, i)

for (key, value) in r.items():

results.setdefault(key, {})

for (ckey, cvalue) in value.items():

results[key].setdefault(ckey, 0)

results[key][ckey] += cvalue

categories = list(results.keys())

categories.sort()

print( "\n\*\*\* Assembling Confusion Matrix ...\n")

print( "\n Classified as: ")

header = " "

subheader = " +"

for category in categories:

header += "% 2s " % category

subheader += "-----+"

print (header)

print (subheader)

total = 0.0

correct = 0.0

for category in categories:

row = " %s |" % category

for c2 in categories:

if c2 in results[category]:

count = results[category][c2]

else:

count = 0

row += " %3i |" % count

total += count

if c2 == category:

correct += count

print(row)

print(subheader)

print("\n%5.3f percent correct" %((correct \* 100) / total))

print("total of %i instances" % total)

# change these to match your directory structure

prefixPath = "C:/Users/karti/Desktop/"

theDir = prefixPath + "dm/"

##theDir = prefixPath ## SHS: otherwise, will throw path error msg

##trainingdir = theDir ## set trainingdir variable to theDir

stoplistfile = prefixPath + "dm.txt"

tenfold( theDir, stoplistfile)

## From the previous file--change these to match your directory structure

## baseDirectory = "/Users/sstreube/Google Drive/ValpoIT633\_DataMining/ValpoIT-533\_DataMiningFall2016/Zacharski/20news-bydate/"

## trainingDir = baseDirectory + "20news-bydate-train/"

## testDir = baseDirectory + "20news-bydate-test/"

## stoplistfile = "/Users/sstreube/Google Drive/ValpoIT633\_DataMining/ValpoIT-533\_DataMiningFall2016/Zacharski/20news-bydate/stopwords0.txt"

## print("Reg stoplist 0 ")

## bT = BayesText(trainingDir, baseDirectory + "stopwords0.txt")

## print("Running Test ...")

## bT.test(testDir)

## print("\n\nReg stoplist 25 ")

## bT = BayesText(trainingDir, baseDirectory + "stopwords25.txt")

## print("Running Test ...")

## bT.test(testDir)

## print("\n\nReg stoplist 174 ")

## bT = BayesText(trainingDir, baseDirectory + "stopwords174.txt")

## print("Running Test ...")

## bT.test(testDir)

Output:  
  
  
%run "C:\Users\karti\Downloads\IT533\_Week\_11TwitterPython\_students(1).py"

Mba @ayu\_surabaya akuuuh naksir sepatumuuuh.

#roommate #malaysia #jakarta #indonesia @ Marc… <https://t.co/IIZU0zDwPj>

Nothing like being treated to a #Starbucks #IcedCoffee by your #roommate on such a #beautifulday to get me through #class 💖

為什麼有人吃東西可以吃的那麼大聲啊🙄🙄#rude #roommate #doesn't #know #how #to #eat #quietly

About to be an epic weekend in dubstep @BukezFinezt @SEP @The\_Widdler @DJYoungsta @Antiserum #jaswan16bitset #vivek… <https://t.co/gvz7fCXFDo>

RT @SomosAnimetaku: ¡Por fin! ✌Acá les tenemos el primer vídeo promocional del anime #RoomMate → <https://t.co/c3JWswxAQ1> 👀<https://t.co/NQ>…

¡Por fin! ✌Acá les tenemos el primer vídeo promocional del anime #RoomMate → <https://t.co/c3JWswxAQ1> 👀<https://t.co/NQj3QUQ7an>

#roommate is majorly #misogynistic 2day. Him &amp;looser pothead bros r debating if it's funny or hilarious to "Hit It &amp; Quit It". #rapeculture

El Anime Room Mate en un Primer Video Promocional

#Anime #RoomMate #NoticiasAnime <https://t.co/gBZSyUmEEU>

Y'all I need a #roommate for #acu21 I want to get to know someone before we are roommates for year. Also nervous to do potluck 😬

Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage... <https://t.co/1StS0r1zha> by… <https://t.co/vRw9tXG4FO>

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @Room\_Mate\_anime: 4月12日放送開始まであと6日！『Room Mate』の最新情報！今日は #前野智昭 さんが演じる #葦原巧 の表情を紹介！柔和な笑顔が素敵！ <https://t.co/wQw2kxwB6Z> #ルームメイト #RoomMate http…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

roommate

In [11]:

Sentiment Analysis on categorized and analyzing :  
  
Input:  
  
  
from \_\_future\_\_ import print\_function

import os, codecs, math

class BayesText:

def \_\_init\_\_(self, trainingdir, stopwordlist, ignoreBucket):

"""This class implements a naive Bayes approach to text

classification

trainingdir is the training data. Each subdirectory of

trainingdir is titled with the name of the classification

category -- those subdirectories in turn contain the text

files for that category.

The stopwordlist is a list of words (one per line) will be

removed before any counting takes place.

"""

self.vocabulary = {}

self.prob = {}

self.totals = {}

self.stopwords = {}

f = open(stopwordlist)

for line in f:

self.stopwords[line.strip()] = 1

f.close()

categories = os.listdir(trainingdir)

#filter out files that are not directories

self.categories = [filename for filename in categories

if os.path.isdir(trainingdir + filename)]

print("\*\*\* init Function: Line 28 Categories and Counting ...\n\n")

for category in self.categories:

print(' ' + category) ## SHS: Verbose

(self.prob[category],

self.totals[category]) = self.train(trainingdir, category,

ignoreBucket)

# I am going to eliminate any word in the vocabulary

# that doesn't occur at least 3 times

toDelete = []

for word in self.vocabulary:

if self.vocabulary[word] < 3:

# mark word for deletion

# can't delete now because you can't delete

# from a list you are currently iterating over

toDelete.append(word)

# now delete

for word in toDelete:

del self.vocabulary[word]

# now compute probabilities

vocabLength = len(self.vocabulary)

print("\*\*\* init Function: Line 48 Computing probabilities:\n\n")

for category in self.categories:

print(' ' + category)

denominator = self.totals[category] + vocabLength

for word in self.vocabulary:

if word in self.prob[category]:

count = self.prob[category][word]

else:

count = 1

self.prob[category][word] = (float(count + 1)

/ denominator)

print ("\*\*\* Line 59 DONE TRAINING\n\n") ## SHS: uncommented verbose

def train(self, trainingdir, category, bucketNumberToIgnore):

"""counts word occurrences for a particular category"""

ignore = "%i" % bucketNumberToIgnore

currentdir = trainingdir + category

directories = os.listdir(currentdir)

print(directories) ## SHS checking variable

counts = {}

total = 0

print("\*\*\* Line 69 START TRAINING FUNCTION ...\n\n")

for directory in directories:

## print(currentdir + '/' + directory) ## SHS: added verbose

if directory != ignore:

## currentBucket = trainingdir + category + "/" + directory

currentBucket = currentdir + "/" + directory ## SHS: Cleaning up variables

files = os.listdir(currentBucket) ## throws error if hidden desktop.ini file is present

## print(" " + currentBucket) ## SHS: verbose

for file in files:

f = codecs.open(currentBucket + '/' + file, 'r', 'iso8859-1')

for line in f:

tokens = line.split()

for token in tokens:

# get rid of punctuation and lowercase token

token = token.strip('\'".,?:-')

token = token.lower()

if token != '' and not token in self.stopwords:

self.vocabulary.setdefault(token, 0)

self.vocabulary[token] += 1

counts.setdefault(token, 0)

counts[token] += 1

total += 1

f.close()

return(counts, total)

def classify(self, filename):

results = {}

for category in self.categories:

results[category] = 0

f = codecs.open(filename, 'r', 'iso8859-1')

print("\*\*\* Line 101: START CLASSIFICATION ...\n\n")

for line in f:

tokens = line.split()

for token in tokens:

print(token)

token = token.strip('\'".,?:-').lower()

if token in self.vocabulary:

for category in self.categories:

if self.prob[category][token] == 0:

print("%s %s" % (category, token))

results[category] += math.log(

self.prob[category][token])

f.close()

results = list(results.items())

results.sort(key=lambda tuple: tuple[1], reverse = True)

# for debugging I can change this to give me the entire list

return results[0][0]

def testCategory(self, direc, category, bucketNumber):

results = {}

directory = direc + ("%i/" % bucketNumber)

print("\*\*\* Line 122 START TESTING the categories now ...\n\n")

print("Testing " + directory)

files = os.listdir(directory)

total = 0

correct = 0

for file in files:

total += 1

result = self.classify(directory + file)

results.setdefault(result, 0)

results[result] += 1

#if result == category:

# correct += 1

return results

def test(self, testdir, bucketNumber):

"""Test all files in the test directory--that directory is

organized into subdirectories--each subdir is a classification

category"""

results = {}

categories = os.listdir(testdir)

#filter out files that are not directories

categories = [filename for filename in categories if

os.path.isdir(testdir + filename)]

correct = 0

total = 0

for category in categories:

#print(".", end="")

results[category] = self.testCategory(

testdir + category + '/', category, bucketNumber)

return results

def tenfold(dataPrefix, stoplist):

results = {}

for i in range(0,10):

bT = BayesText(dataPrefix, stoplist, i)

r = bT.test(theDir, i)

for (key, value) in r.items():

results.setdefault(key, {})

for (ckey, cvalue) in value.items():

results[key].setdefault(ckey, 0)

results[key][ckey] += cvalue

categories = list(results.keys())

categories.sort()

print( "\n\*\*\* Assembling Confusion Matrix ...\n")

print( "\n Classified as: ")

header = " "

subheader = " +"

for category in categories:

header += "% 2s " % category

subheader += "-----+"

print (header)

print (subheader)

total = 0.0

correct = 0.0

for category in categories:

row = " %s |" % category

for c2 in categories:

if c2 in results[category]:

count = results[category][c2]

else:

count = 0

row += " %3i |" % count

total += count

if c2 == category:

correct += count

print(row)

print(subheader)

print("\n%5.3f percent correct" %((correct \* 100) / total))

print("total of %i instances" % total)

# change these to match your directory structure

prefixPath = "C:/Users/karti/Desktop/"

theDir = prefixPath + "dm/"

##theDir = prefixPath ## SHS: otherwise, will throw path error msg

##trainingdir = theDir ## set trainingdir variable to theDir

stoplistfile = prefixPath + "dm.txt"

tenfold( theDir, stoplistfile)

## From the previous file--change these to match your directory structure

## baseDirectory = "/Users/sstreube/Google Drive/ValpoIT633\_DataMining/ValpoIT-533\_DataMiningFall2016/Zacharski/20news-bydate/"

## trainingDir = baseDirectory + "20news-bydate-train/"

## testDir = baseDirectory + "20news-bydate-test/"

## stoplistfile = "/Users/sstreube/Google Drive/ValpoIT633\_DataMining/ValpoIT-533\_DataMiningFall2016/Zacharski/20news-bydate/stopwords0.txt"

## print("Reg stoplist 0 ")

## bT = BayesText(trainingDir, baseDirectory + "stopwords0.txt")

## print("Running Test ...")

## bT.test(testDir)

## print("\n\nReg stoplist 25 ")

## bT = BayesText(trainingDir, baseDirectory + "stopwords25.txt")

## print("Running Test ...")

## bT.test(testDir)

## print("\n\nReg stoplist 174 ")

## bT = BayesText(trainingDir, baseDirectory + "stopwords174.txt")

## print("Running Test ...")

## bT.test(testDir)

**File output:**Mba @ayu\_surabaya akuuuh naksir sepatumuuuh.

#roommate #malaysia #jakarta #indonesia @ Marc… https://t.co/IIZU0zDwPj

Nothing like being treated to a #Starbucks #IcedCoffee by your #roommate on such a #beautifulday to get me through #class 💖

為什麼有人吃東西可以吃的那麼大聲啊🙄🙄#rude #roommate #doesn't #know #how #to #eat #quietly

About to be an epic weekend in dubstep @BukezFinezt @SEP @The\_Widdler @DJYoungsta @Antiserum #jaswan16bitset #vivek… https://t.co/gvz7fCXFDo

RT @SomosAnimetaku: ¡Por fin! ✌ Acá les tenemos el primer vídeo promocional del anime #RoomMate → https://t.co/c3JWswxAQ1 👀 https://t.co/NQ…

¡Por fin! ✌ Acá les tenemos el primer vídeo promocional del anime #RoomMate → https://t.co/c3JWswxAQ1 👀 https://t.co/NQj3QUQ7an

#roommate is majorly #misogynistic 2day. Him &amp;looser pothead bros r debating if it's funny or hilarious to "Hit It &amp; Quit It". #rapeculture

El Anime Room Mate en un Primer Video Promocional

#Anime #RoomMate #NoticiasAnime https://t.co/gBZSyUmEEU

Y'all I need a #roommate for #acu21 I want to get to know someone before we are roommates for year. Also nervous to do potluck 😬

Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage... https://t.co/1StS0r1zha by… https://t.co/vRw9tXG4FO

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @Room\_Mate\_anime: 4月12日放送開始まであと6日！『Room Mate』の最新情報！今日は #前野智昭 さんが演じる #葦原巧 の表情を紹介！柔和な笑顔が素敵！ https://t.co/wQw2kxwB6Z #ルームメイト #RoomMate http…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

RT @AskAprilcom: Dear #AskApril,

Our #roommate broke up with a #girlfriend &amp; wants to #sabotage my #relationship with my #boyfriend. https:…

roommate