```
import json
import os
import sqlite3
import time
t0=time.time()

db = sqlite3.connect('data/thdb')
cursor = db.cursor()

cursor.execute('''CREATE TABLE IF NOT EXISTS tweets(id INTEGER PRIMARY KEY, tid INT, tweet TEXT, tweetNP TEXT, tweetNPSW TEXT, class TEXT)''')
cursor.execute('''CREATE TABLE IF NOT EXISTS words(id INTEGER PRIMARY KEY, word TEXT, count INTEGER NOT NULL DEFAULT 0, hCount INTEGER NOT NULL DEFAULT 0, tCount INTEGER NOT NULL DEFAULT 0, UNIQUE(word))''')
cursor.execute('''CREATE TABLE IF NOT EXISTS wordsp(id INTEGER PRIMARY KEY, word TEXT, phil FLOAT, ptru FLOAT, UNIQUE(word))''')
db.commit()

punc = '!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~’…—'
stopWordsNoP = ["rt","a","about","above","after","again","against","all","am","an","and","any","are","arent","as","at","be","because","been","before","being","below","between","both","but","by","cant","cannot","could","couldnt","did","didnt","do","does","doesnt","doing","dont","down","during","each","few","for","from","further","had","hadnt","has","hasnt","have","havent","having","he","hed","hell","hes","her","here","heres","hers","herself","him","himself","his","how","hows","i","id","ill","im","ive","if","in","into","is","isnt","it","its","its","itself","lets","me","more","most","mustnt","my","myself","no","nor","not","of","off","on","once","only","or","other","ought","our","ours","ourselves","out","over","own","same","shant","she","shed","shell","shes","should","shouldnt","so","some","such","than","that","thats","the","their","theirs","them","themselves","then","there","theres","these","they","theyd","theyll","theyre","theyve","this","those","through","to","too","under","until","up","very","was","wasnt","we","wed","well","were","weve","were","werent","what","whats","when","whens","where","wheres","which","while","who","whos","whom","why","whys","with","wont","would","wouldnt","you","youd","youll","youre","youve","your","yours","yourself","yourselves"]

def stopWordRemover (text, swords):
    return list(set([word for word in text.split() if word.lower() not in swords]))

def linkRemover (text):
    return list(set([word for word in text.split() if not (word.lower().startswith('www.') or word.lower().startswith('http'))]))

for jsonFilename in os.listdir('dumpKaggle'):
    with open("dumpKaggle\\"+jsonFilename) as json_file:
        data = json.load(json_file)
        tid = data["id"]
        tweet = data["text"].lower()
        classOf = data["handle"]
        tweetNP = tweet.translate(str.maketrans("","", punc))
        tweetNPSW = ' '.join(stopWordRemover(tweetNP, stopWordsNoP))
        tweetNPSW = ' '.join(linkRemover(tweetNPSW))

        cursor.execute('''INSERT INTO tweets(tid, tweet, tweetNP, tweetNPSW, class) VALUES(?,?,?,?,?)''', (tid, tweet, tweetNP, tweetNPSW, classOf))

        words = tweetNPSW.split()
        for word in words:
            cursor.execute('''INSERT OR IGNORE INTO words(word) VALUES(?)''', (word,))
            cursor.execute('''UPDATE words SET count = count + ? WHERE word = ?''', (1, word))
            if classOf == "HillaryClinton":
                cursor.execute('''UPDATE words SET hCount = hCount + ? WHERE word = ?''', (1, word))
            elif classOf == "realDonaldTrump":
                #classOf = "disliked"
                cursor.execute('''UPDATE words SET tCount = tCount + ? WHERE word = ?''', (1, word))
            else:
                print(classOf)

        db.commit()

cursor.execute('''SELECT class from tweets''')
tweets = cursor.fetchall()
tweetCount = float(len(tweets))
HillaryCount = float(tweets.count(('HillaryClinton',)))
TrumpCount = float(tweets.count(('realDonaldTrump',)))
tHillaryCount = float(tweets.count(('HillaryClinton',)))/tweetCount # p(hillar)
tTrumpCount = float(tweets.count(('realDonaldTrump',)))/tweetCount # p(trump)

print('There are '+str(tweetCount)+' tweets in the database\n'+str(HillaryCount)+" %"+str(tHillaryCount)+' labeled as Hillary\n'+str(TrumpCount)+" %"+str(tTrumpCount)+' labeled as Trump\n')

cursor.execute('''SELECT word FROM words''')
wordCount = float(len(cursor.fetchall()))
print('There are '+str(wordCount)+' words in the database')

cursor.execute('''SELECT hCount from words''')
hCount = cursor.fetchall()
hCount = float(len(hCount) - hCount.count((0,)))
print('There are '+str(hCount)+' Hillary labeled words in the database')

cursor.execute('''SELECT tCount from words''')
tCount = cursor.fetchall()
tCount = float(len(tCount) - tCount.count((0,)))
print('There are '+str(tCount)+' Trump labeled words in the database')

cursor.execute('''SELECT sum(hCount) FROM words''')
hWordOccurances = float(cursor.fetchall()[0][0])

cursor.execute('''SELECT sum(tCount) FROM words''')
tWordOccurances = float(cursor.fetchall()[0][0])

print(str(hWordOccurances)+' words in Hillary tweets\n'+str(tWordOccurances
)+' words in Trump tweets\n')

cursor.execute('''SELECT * from words''')
rows = cursor.fetchall()
for row in rows:
    cursor.execute('''INSERT OR IGNORE INTO wordsp(word) VALUES(?)''', (row[1],))
    phil = (float(row[3])+1.0)/(hWordOccurances+wordCount)*tHillaryCount
    ptru = (float(row[4])+1.0)/(tWordOccurances+wordCount)*tTrumpCount
    cursor.execute('''UPDATE wordsp SET phil = ? WHERE word = ?''', (phil, row[1]))
    cursor.execute('''UPDATE wordsp SET ptru = ? WHERE word = ?''', (ptru, row[1]))
    db.commit()

db.close()
t1 = time.time()
print("Total time: "+str(t1-t0))
print("Learning finished")
```