Mids3

import pandas as pd

trainData = pd.read\_csv("train.csv")

testData = pd.read\_csv("test.csv")

trainData.sample(frac=1).head(5)

from sklearn.feature\_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(min\_df = 5,

max\_df = 0.8,

sublinear\_tf = True,

use\_idf = True)

train\_vectors = vectorizer.fit\_transform(trainData['Content'])

test\_vectors = vectorizer.transform(testData['Content'])

import time

from sklearn import svm

from sklearn.metrics import classification\_report

classifier\_linear = svm.SVC(kernel='linear')

t0 = time.time()

classifier\_linear.fit(train\_vectors, trainData['Label'])

t1 = time.time()

prediction\_linear = classifier\_linear.predict(test\_vectors)

t2 = time.time()

time\_linear\_train = t1-t0

time\_linear\_predict = t2-t1

print("Training time: %fs; Prediction time: %fs" % (time\_linear\_train, time\_linear\_predict))

report = classification\_report(testData['Label'], prediction\_linear, output\_dict=True)

print('positive: ', report['pos'])

print('negative: ', report['neg'])