學號:B0928031

姓名:鄭茹云

學系:人工智慧三甲

連結:https://colab.research.google.com/drive/1-Cmwt5E8YsSu4I7nwSt9XCeqOjkhZAnT?usp=share_link

```
import  os
print(os.getcwd())
```

```
C:\Users\layla\Downloads
```

```
print(os.path.abspath("hw2.json"))
```

```
C:\Users\layla\Downloads\hw2.json
```

```
import  json
from  sklearn.feature_extraction.text  import  TfidfVectorizer
from  sklearn.neighbors  import  KNeighborsClassifier
from  sklearn.svm  import  SVC
from  sklearn.metrics  import  accuracy_score

#  讀取  hw2.json  檔
file  =  open('hw2.json',  'r',  encoding='utf-8')

mm=[]
#會出現error因為json內容太大，一行一行讀取在合併

for  line  in  file.readlines():
        movie_data  =  json.loads(line)
        mm.append(movie_data)

X  =  []
y  =  []
for  movie  in  mm:
        if  movie['label']  !=  'NA':  #忽略NA
                X.append(movie["intro"])
                y.append(movie['label'])




vectorizer  =  TfidfVectorizer()
X_tfidf  =  vectorizer.fit_transform(X)

#500拿來評估
train_X_tfidf  =  X_tfidf[:-500]
train_y  =  y[:-500]
test_X_tfidf  =  X_tfidf[-500:]
test_y  =  y[-500:]
```

```
knn_classifier = KNeighborsClassifier(n_neighbors=5)
knn_classifier.fit(train_X_tfidf, train_y)
knn_y_pred = knn_classifier.predict(test_X_tfidf)
knn_accuracy = accuracy_score(test_y, knn_y_pred)

print('KNN classifier accuracy: %.2f%%' % (knn_accuracy * 100))
```

```
KNN classifier accuracy: 38.80%
C:\Users\layla\anaconda3\lib\site-packages\sklearn\neighbors\_classification.py:228: FutureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the
  mode, _ = stats.mode(_y[neigh_ind, k], axis=1)
```

```
# SVM
svm_classifier = SVC(kernel='linear')
svm_classifier.fit(train_X_tfidf, train_y)
svm_y_pred = svm_classifier.predict(test_X_tfidf)
svm_accuracy = accuracy_score(test_y, svm_y_pred)


print('SVM classifier accuracy: %.2f%%' % (svm_accuracy * 100))
```

```
SVM classifier accuracy: 38.40%
```