

IRTM HW3 report

B11705058 資管三 廖振翔

執行環境: vscode

程式語言: python 3.11.5

執行方式: 在與程式的同個目錄下放入 training.txt 還有 test_docs 資料夾，其中 test_docs 資料夾中放入所有的文檔，執行 python 程式碼即可

pip install 套件: 要先 install nltk 才能使用以下套件

```
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
```

作業處理邏輯說明: 先讀取所有資料後 stemmer 跟 remove stopword，再計算詞頻和文檔頻率，使用 chi-square 選擇特徵，接著訓練 Naive Bayes classifier，最後 predict 並輸出為 submission.csv