**爬文章.py**

42行 將想要爬的ptt版的網址

202行 將要貯存的位置輸入

204行 將想要的檔案格式更改如txt

**整理成文章.py**

16行 輸入文章所在檔案位置

45~58行將要的文章內容留下

71~89行將要的評論內容留下

100行選擇寫入的檔案

**分離文章.py**

將同個作者的文章分離

3行輸入文章作者及完成

**匯入ela.py**

14行輸入檔案位置

21行 輸入index type

**搜尋ela裡的文章.py**

11行~15行輸入欲查詢之內容之數量

20行 輸入寫入文章

**整理統計.py**

6行輸入檔案名

7行統計欲統計之數量

**Tokenlize**

86-102 讀訓練集和測試集的文檔並用結巴切好在存到另一個資料夾

**Evaluate**

41-47計算文檔的tf-idf

54-59 依所選來選擇演算法，這邊的參數是手動調但可以使用gridsearchCV

metrics\_result輸出測試集的準確率

70-83結巴切要預測的文章，計算tf-idf，預測並輸出結果。