**災難推文辨識**

**小組成員和負責工作：**

**109590041 范遠皓 撰寫程式和測試**

**109590043 柯瑞霖 撰寫程式和報告**

**環境**

使用的語言：Python

所需套件：

requirements.txt 檔案內

[pytorch安裝](https://pytorch.org/get-started/locally/)

安裝辦法：

Zip檔內有 requirements.txt

終端機輸入以下指令安裝

pip3 install -r requirements.txt

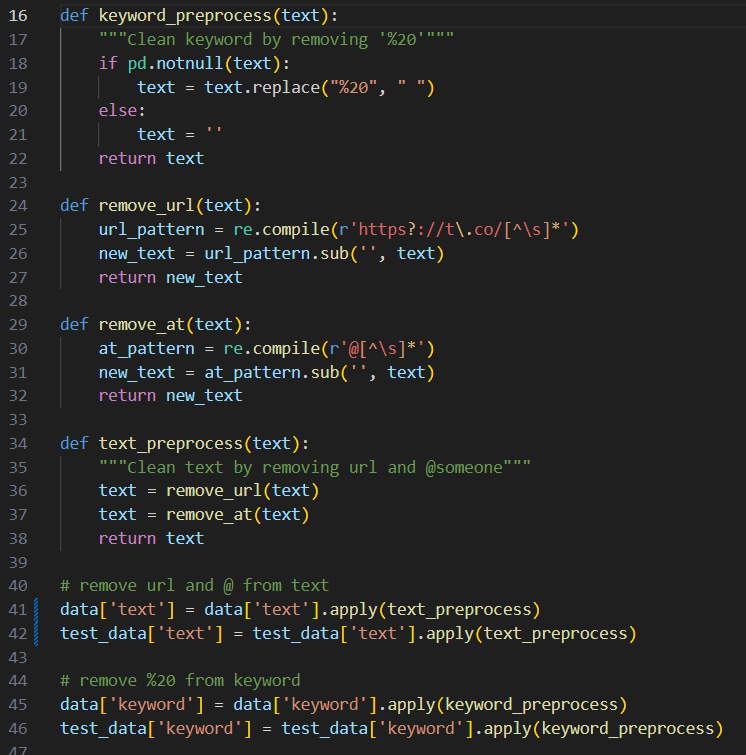
**資料讀取 :**

正在插入影像...

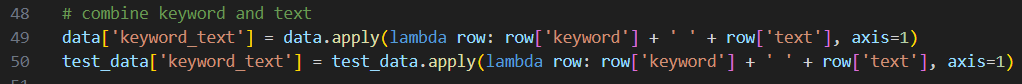
**資料預處理 :**

將 keyword 標籤內有包含 '%20' 的替換掉

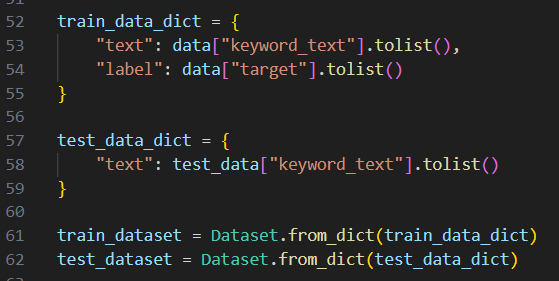
移除網址 、 @someone



將 keyword 加入 text(tweet)

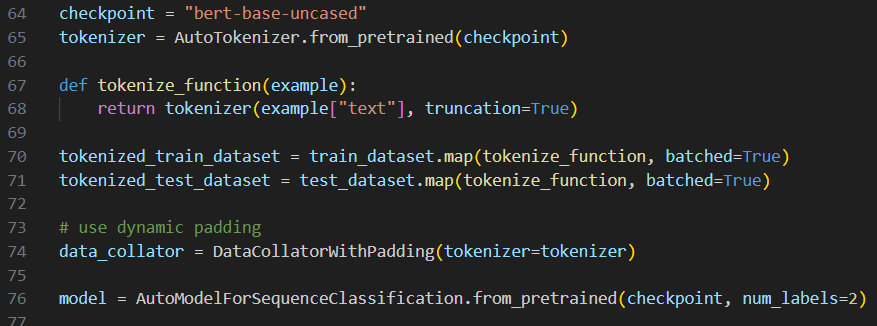


將資料轉成 dataset

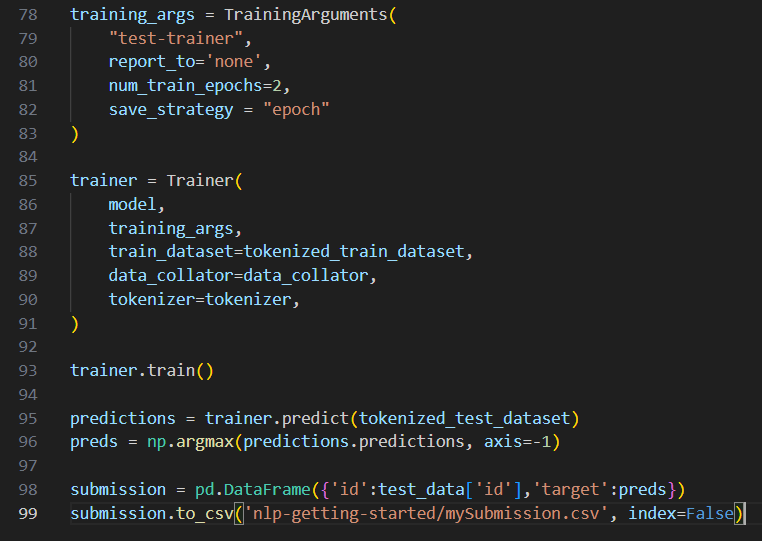


**載入BERT模型 :**

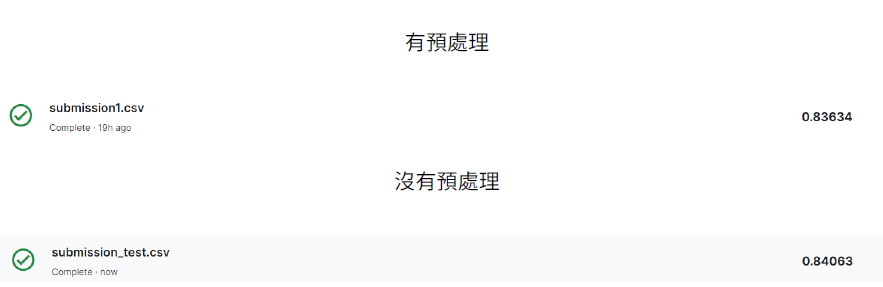
初始化一個 BERT tokenizer ，並將文本轉換成 BERT 模型所需的輸入格式，載入預訓練模型BERT，用於訓練一個基於BERT模型的序列分類器



**調整訓練參數並輸出結果 :**



**執行結果：**



**找出受災地區 :**

使用了spaCy套件的en\_core\_web\_sm模型，檢測每條貼文中的地點

