Programming Assignment 3 作業說明

Task

本次要做航空公司評論的 **Sentiment Analysis**,請根據 data.csv 訓練與建立模型,並預測 predict.csv 中評論的情緒標記結果,將結果輸出為 result.csv 繳交。

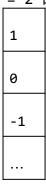
Dataset

本次作業提供以下兩個檔案,共有 36860 筆資料:

- data.csv:作為模型訓練使用,共有 31860 筆,包含以下兩個欄位。
 - content:針對航空公司提供的服務做出的文字評論。
 - category:標籤評論是屬於負面 (-1)、中立 (0) 或 正面(1)態度。
- predict.csv:預測結果使用,共 5000 筆 (無標記),包含以下兩個欄位。
 - o num:每行評論的編號。
 - content:針對航空公司提供的服務做出的文字評論。
- * 若使用 pandas 預設用 utf-8 編碼在讀取 csv 檔案時會有些字元會遇到問題,可以改用 latin-1 編碼讀取。

Grading

- 1. report and code (40%): report 需詳細說明使用的模型架構與邏輯,約 1~2頁; code 部分則須繳交完整的程式碼。
- 2. result (60%): 將 predict.csv 預測結果輸出成 result.csv,格式如下
 - 輸出一個長度為 5000 的 column,每行對應評論的預測結果,只可能出現 1/0/-1 三種值。以下表為例,第一行的 1 為 num = 1 的評論預測結果,第二行的 0 為 num = 2 的評論預測結果,以此類推。



*輸出結果不包含header與index。

Notice

- 1. 繳交期限:5/23 (Tue.) 23:59,不接受遲交。
- 2. 需從頭訓練一個模型,<mark>不可以</mark>使用 pre-trained model (e.g., BERT) 。
- 3. 若 result.csv 格式錯誤,則作業以 0 分計算。但有一次補交的機會,補交 後分數會以九折計算。