作業6: 眾裡尋他千百度,那模型就在燈火欄珊處

現在我們就要開始來使用江湖中傳說的文字探勘絕學啦!

為了確保大家的學習品質,因此課程中一半的內容將做為作業的一部份;另一部份才是讓大家帶回去練習的作業唷!那我們就開始吧!!

1. 請問啟spam.xls檔案! 並開始進行分析,開始前請針對需要的欄位進行必要之前置處理. (10%)

2. 接下來,我們要針對文字進行分析,因此有一些必要的處理程序,請說明一下您的步驟(10%),並請說明下列分析結果的意涵: TF-IDF (5%)

3. 接下來,我們要開發製做一個垃圾郵件過濾器,請找出您能找出正確率最高的分類機制,並略加說明(15%)

\*註:因為有正確率評估的關係,請記得要使用訓練/測試集,又或是Cross-Validation

4. 接下來,承上述步驟,請使用Womens Clothing E-Commerce Reviews.csv檔案進行分析,在這邊我們希望您可以借由分析評論來預測評分或是您認為值得分析的議題. 另外也請您自行思考一下還有什麼分析可以進行並說明結果唷 (60%)。