1. 執行環境：jupyer notebook
2. 程式語言：python 3.11.9
3. 執行方式：os：處理文件與目錄操作。pandas：用於數據處理和操作。nltk：進行文本處理，包括去除停用詞和詞幹提取。re：執行正則表達式處理文本。math 和 numpy：用於數學運算。
4. 作業邏輯說明：總共分成三個區塊：第一個區塊是數據加載與準備：從指定資料夾 IRTM 中加載所有文檔，按文件名排序後存入 corpus。從 training.txt 中加載文檔與類別對應關係，生成一個字典 labels，用於標記訓練數據。。
   第二個區塊是數據處理：將文檔分為訓練集和測試集：
   訓練集：根據 training.txt 提供的標籤提取已標記的文檔。
   測試集：未在 training.txt 中標記的文檔被歸為測試集，類別設為空值。使用文本處理技術（正則表達式清理、去停用詞、詞幹提取等）對文檔進行預處理，並為每篇文檔計算詞頻（TF）。第三個區塊是特徵選擇：使用卡方檢驗（Chi-Square）方法選擇高信息量的詞作為特徵，過濾掉冗餘詞，最終選取前 300 個詞作為模型的特徵。(題目規定 500 內，但發現少一點之後效果比較好)
   最後一個區塊是模型訓練與應用：使用 Naive Bayes 分類器模型進行訓練，根據詞頻和選擇的特徵計算類別的先驗概率和條件概率。測試集應用訓練好的模型，基於測試文檔的詞頻向量計算類別概率，並將文檔分到最大概率的類別。