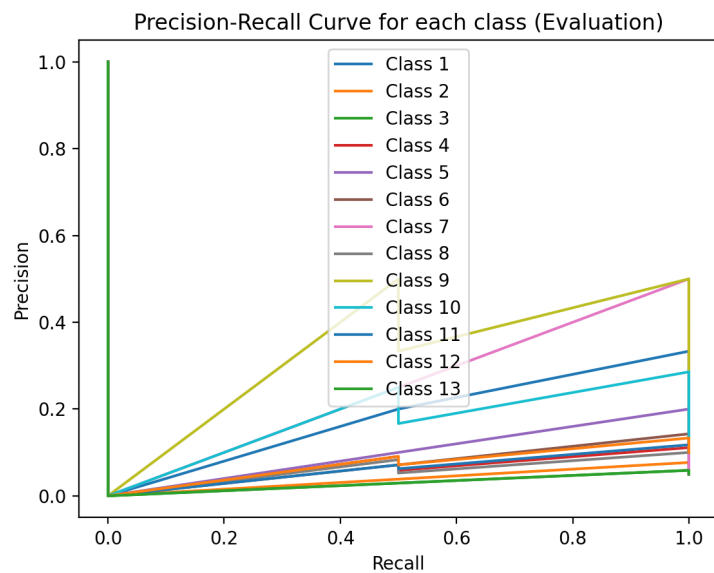


## Text Mining HW2 Report R12723006 財金碩一 林姝延

- 執行環境：vs code
- 執行語言：python 3.12.0
- 執行方式：將存有 1095 個 txt.檔的 PA1-data 資料夾以及 training\_new.txt 檔案存放在與當前執行程式相同的路徑下，使用 os 套件取得當前路徑，並直接讀取同樣位於此路徑的資料夾與檔案。而最終輸出的分類結果 csv 檔案同樣會存到與程式碼相同路徑的位置。
- 邏輯說明：首先讀取 1095 個 txt 檔案，建構一個 data frame，並讀取 training\_new.txt 檔案，取得各個 class 作為 training dataset 的文章編號，另將 training dataset 中 10% 設為 evaluation set，並把 data frame 中 training 文章刪除，成為 testing dataset。接著建構 multi-hot vectors 或是 tfidf vectors，並進行訓練及預測。最終計算並呈現各種方法的 precision, recall, f1 結果，並畫出 precision-recall curve。
- 執行結果：  
使用各方法所得的各組 scores 以及 precision-recall curve 如下：

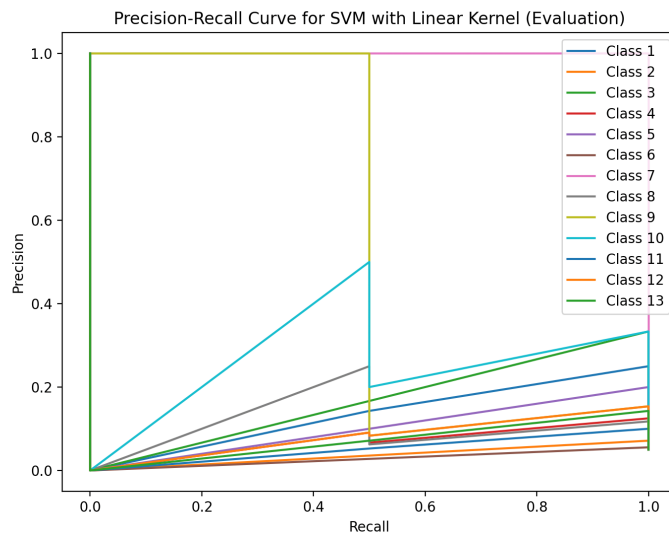
- Bernoulli Naïve Bayes method

	precision	recall	f1-score
1	0.00	0.00	0.00
2	0.00	0.00	0.00
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	0.00	0.00	0.00
6	0.08	1.00	0.15
7	0.50	1.00	0.67
8	0.00	0.00	0.00
9	0.00	0.00	0.00
10	0.00	0.00	0.00
11	0.00	0.00	0.00
12	0.00	0.00	0.00
13	0.00	0.00	0.00
accuracy			0.10
macro avg	0.04	0.15	0.06
weighted avg	0.03	0.10	0.04



- SVM method, Linear kernel

	precision	recall	f1-score
1	0.33	1.00	0.50
2	0.50	0.33	0.40
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	0.00	0.00	0.00
6	0.00	0.00	0.00
7	0.00	0.00	0.00
8	0.00	0.00	0.00
9	0.00	0.00	0.00
10	0.00	0.00	0.00
11	1.00	0.33	0.50
13	0.00	0.00	0.00
accuracy			0.15
macro avg	0.15	0.14	0.12
weighted avg	0.24	0.15	0.16



- SVM method, RBF kernel

	precision	recall	f1-score	support
1	0.00	0.00	0.00	2
2	0.00	0.00	0.00	1
3	0.00	0.00	0.00	1
4	0.00	0.00	0.00	2
5	0.00	0.00	0.00	1
6	0.00	0.00	0.00	1
7	1.00	1.00	1.00	1
8	0.00	0.00	0.00	2
9	0.50	0.50	0.50	2
10	1.00	0.50	0.67	2
11	0.00	0.00	0.00	2
12	0.00	0.00	0.00	2
13	0.00	0.00	0.00	1
accuracy			0.15	20
macro avg	0.19	0.15	0.17	20

