Kaggle 個人賽實驗報告 Kaggle Individual Competition Report

*[姓名 Name]*

*[學號 Student ID]*

*[系別單位 Department]*

# 資料集前處理 Dataset Preprocessing

*[說明如何處理資料 Briefly explain your data processing methodology]*

# 實驗紀錄 Experiments

1. Machine Learning Classifiers

Table 1. Experiment Records of basic classificers

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Methods | Training Data k-fold  Average-Micro  Precision/Recall/F1 | Training Data k-fold  Average-Macro  Precision/Recall/F1 | Test Data  Public Leaderboard  Score/Rank/filename | Test Data  Private Leaderboard  Score/Rank/filename | Parameters for training model |
| Naïve Bayes #1 | 0.9999/0.9999/0.9999 | 0.9999/0.9999/0.9999 | 0.99/1/naivebayes.csv | 0.99/1/naivebayes.csv | MultinomialNB(alpha=1.0)  TfidfVectorizer(max\_df=0.9, min\_df=0.001, ngram\_range=(1, 2), preprocessor=<function dummy\_fun at 0x7f5e9bf77600>, token\_pattern=None, tokenizer=<function dummy\_fun at 0x7f5e9bf77600>) |
| Naïve Bayes #2 |  |  |  |  |  |
| Naïve Bayes #3 |  |  |  |  |  |
| Logistic Regression #1 |  |  |  |  |  |
| Logistic Regression #2 |  |  |  |  |  |
| Logistic Regression #3 |  |  |  |  |  |
| Decision Tree #1 |  |  |  |  |  |
| Decision Tree #2 |  |  |  |  |  |
| Decision Tree #3 |  |  |  |  |  |
| KNN #1 |  |  |  |  |  |
| KNN #2 |  |  |  |  |  |
| KNN #3 |  |  |  |  |  |
| SVM #1 |  |  |  |  |  |
| SVM #2 |  |  |  |  |  |
| SVM #3 |  |  |  |  |  |

[最低要求是實作Naïve Bayes, Logistic Regression, Decision Tree, KNN, SVM 等五種基礎分類方法，並紀錄訓練模型使用的參數，以 K-folds 訓練跑分 Average-Micro 及Average-Macro 的Precision/Recall/F1 表現，最後用訓練好的方法去推論測試資料集的答案並上傳 Kaggle 獲得 Public/Private排行榜的分數及排名。每種方法至少要有 3 次上傳預測檔成功紀錄，登錄上傳檔名。

Document experiments using five basic classification methods discussed in class: Naïve Bayes, Logistic Regression, Decision Tree, KNN, and SVM. Include:

- Training parameters

- Model performance using K-folds cross-validation:

- Average-Micro Precision/Recall/F1 scores

- Average-Macro Precision/Recall/F1 scores

- Kaggle submission results:

- Test set predictions

- Public/Private leaderboard scores and rankings

At least 3 successful submission records per method with submission filenames]

1. 你的方法 Your Methods

Table 2. Experiment Records of your methods

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Methods | Training Data k-fold  Average-Micro  Precision/Recall/F1 | Training Data k-fold  Average-Macro  Precision/Recall/F1 | Test Data  Public Leaderboard  Score/Rank/filename | Test Data  Private Leaderboard  Score/Rank/filename/Date | Parameters for training model |
| Your method #1 |  |  |  |  |  |
| Your method #2 |  |  |  |  |  |
| Your method #3 |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

*[實驗你的方法，紀錄訓練參數，訓練模型以 K-folds 跑分 Average-Micro 及Average-Macro 的Precision/Recall/F1 表現，最後以訓練好方法去推論測試資料集的答案並上傳 Kaggle 獲得 Public/Private排行榜的分數及排名。每種方法至少要有 3 次上傳預測檔成功紀錄，登錄上傳檔名。*

*Document experiments using your methods. Include:*

*- Training parameters*

*- Model performance using K-folds cross-validation:*

*- Average-Micro Precision/Recall/F1 scores*

*- Average-Macro Precision/Recall/F1 scores*

*- Kaggle submission results:*

*- Test set predictions*

*- Public/Private leaderboard scores and rankings*

*At least 3 successful submission records per method with submission filenames]*

1. 討論 Discussion

*[討論前處理、模型方法及參數調整的利弊得失。[Provide a concise analysis of:*

*- Preprocessing effectiveness*

*- Model performance comparison*

*- Parameter tuning trade-offs and impacts]]*