資料探勘與大數據

期末專題

期末專題內容

1. **主題：**針對自己組別抽到的主題進行描述型、診斷型，與預測型資料分析
2. **預測類型**：分類或回歸
3. **預測方法：課堂**學過或其他自學的方法
4. **特徵：**不限，可將所含的基本特徵組合成其他具有預測力的衍生特徵。
5. **模型評估方式：**將資料集分為訓練資料與測試資料，訓練資料用於建立模型，測試資料用於評估模型準確率。訓練與測試資料的切割比例應約介於8:2到6:4之間。
6. **評估指標：**Precision、Recall、F1-Score等
7. **報告內容:**

報告內容應包含下列要素：

1. 研究主題與動機描述
2. 方法介紹
3. 特徵的設計與設計原理
4. 特徵、預測目標的敘述性統計(可用excel計算)，並對結果加以描述
5. 模型的設計(模型參數、訓練驗證比例等)
6. 訓練資料、測試資料的Precision、Recall、F1-Score結果
7. 特徵重要性呈現，並以理論解釋結果
8. 結論
9. (非必要)其他額外分析的內容(如資料視覺化、不同模型的比較、不平衡資料判斷與解決方法等)

Datasets (括號中的數字代表可選該資料集的組別數):

1. Lending club loans dataset(2) <https://www.kaggle.com/wordsforthewise/lending-club>
2. default of credit card clients dataset(2) <https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>
3. 南部某醫院餐廳商品銷售量(2)
4. KK Box Churn Prediction(2)

<https://www.kaggle.com/c/kkbox-churn-prediction-challenge>