

Data Mining Homework2 作業模板

作業回答共 20 分，未繳交或未作答 0 分計算。

1. 方法選擇 (Method Selection) 4 分

- 請說明你選擇的模型，及為什麼選擇這個方法？為什麼不選擇其他方法嗎？

2. 特徵工程 (Feature Engineering) 8 分

- 你有處理缺失值 (Missing Values) 嗎？
- 你有做特徵篩選 (Feature Selection) 嗎？使用了什麼方法？
- 是否有進行特徵轉換 (例如標準化、偏度處理)？

3. 交叉驗證與模型調整 (Cross-validation & Model Tuning) 5 分

- 你使用了哪種交叉驗證方法 (k-fold, stratified k-fold, train-test split, etc.)？
- 如何選擇最佳超參數？是否使用 learning curve / grid search / random search？
- 請附上 learning curve 或 tuning 過程的圖表。

4. 模型表現與分析 (Model Performance & Analysis) 3 分

- 請畫出混淆矩陣 (confusion matrix)
- 你的最終模型表現如何 (Test AUC, Valid AUC etc.)？D
- (可選) 你有比較不同方法的表現嗎？如果有，結果如何？

5. 結論與反思 (Conclusion & Reflection)

- 這次作業中遇到的主要挑戰是什麼？
- 你認為你的模型還可以如何改進？
-