**程式設計期末專題 – 分類**

**組員: 范睿凱、謝宗穎**

**分析原因:**

首先，我們想了解鐵達尼號生還者與其他因素的關係，而決定利用決策樹這個機器學習方法。原因主要是它有不少優點: 簡單且高度的可解釋性、低計算時間複雜度、決策階段都明確、較少要調整的超參數。第二，我們想展現這學期學到的程式技巧與內容，並利用分類的過程與結果來展現成果。最後，判斷準確度是否如預期。

**程式碼的過程:**

1. **導入標準庫:**利用import 的方式，從python內建的標準函式庫導入。例如:pandas、DecisionTreeClassifier等等。
2. **匯入檔案:** 把train.csv 檔匯入，並把不重要的因素 – PassengerId及要預測的項目Survived，利用drop 刪掉。
3. **處理缺失值:** 我們得到的數據有些是同質的，且有可能會丟失，所以需要對其進行處理，以免降低機器學習模型的性能。
4. **減少特徵數**: 再減少不必要的特徵，讓機器學習效率增加。
5. **建立模型**:也就是將數據集拆分為訓練集和測試集，並加入參數增加準確度。
6. **預測結果及準確度:** 將預測的結果及準確度印出來。