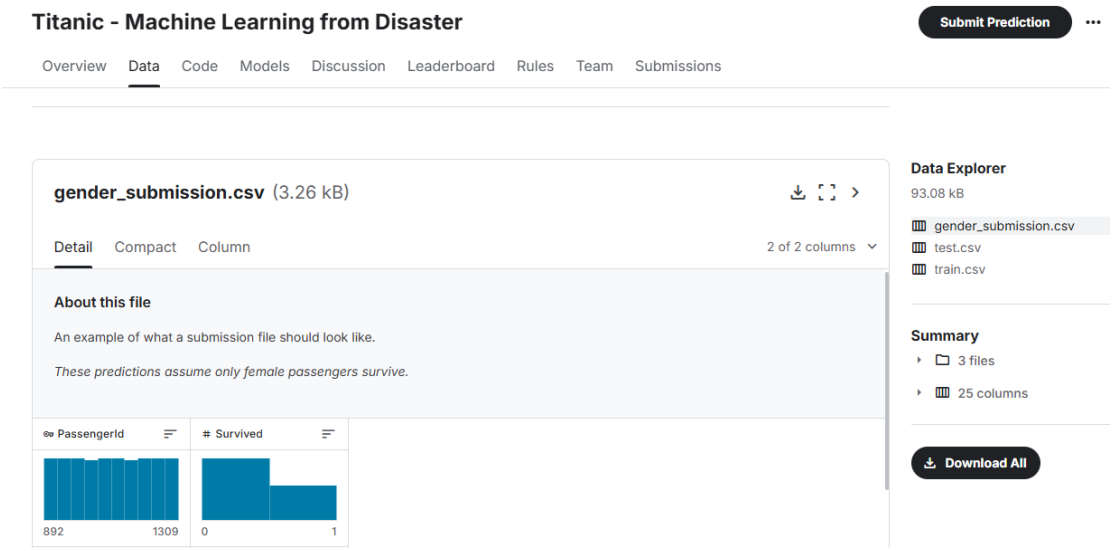
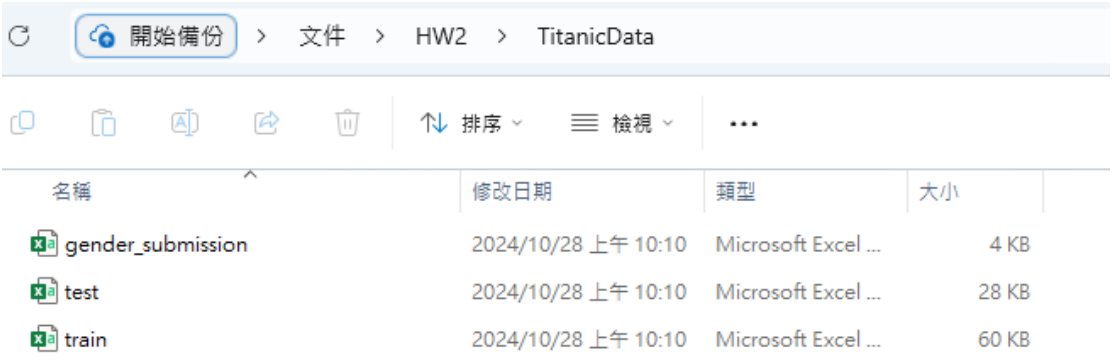


使用的 Data 是 kaggle 內 Titanic - Machine Learning from Disaster 下載的 主要用到的是 test.csv 和 train.csv



使用的是 vs code 中 Jupyter Notebook (.ipynb)的環境

下面是作業要求的 confusion matrix 還有 accuracy

準確率在 82%附近

```
    with torch.no_grad():
        y_true = valid_y.numpy()  # 驗證集的真實標籤
        y_pred = best_model(valid_x).round().numpy()  # 驗證集的預測結果

        matrix = confusion_matrix(y_true, y_pred)
        print("Confusion Matrix:")
        print(matrix)
```

[522] ✓ 0.0s

```
Confusion Matrix:
[[107  10]
 [ 21  41]]
```

```
    with torch.no_grad():
          n_sample=valid_x.shape[0]
          pre=best_model(valid_x)
          pre=pre.round()
          n_correct=(pre==valid_y).sum()
          acc=n_correct/n_sample
          print(f'valid_acc={acc:.4f}'  )
```

[516] ✓ 0.0s

```
valid_acc=0.8268
```

```
    with torch.no_grad():
          n_sample=train_x.shape[0]
          pre=best_model(train_x)
          pre=pre.round()
          n_correct=(pre==train_y).sum()
          acc=n_correct/n_sample
          print(f'train_acc={acc:.4f}'  )
```

[517] ✓ 0.0s

```
train_acc=0.8216
```

確認最後生成的 IT_submission.csv 格式正確

418 筆資料

Passenger ID 沒有重複

Survived 欄位的唯一值為 [0, 1]

HW2.ipynb ● 1.py ×

1.py > ...

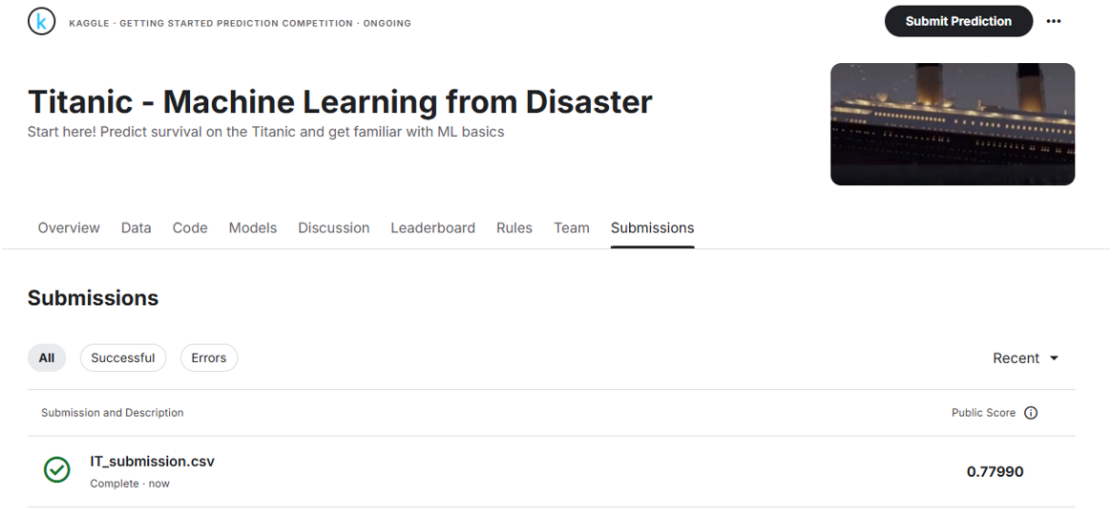
```python
import pandas as pd

# 載入 CSV 文件
df = pd.read_csv('IT_submission.csv')

# 顯示前幾筆資料
print(df.head())

# 檢查資料的資訊
print(df.info())

# 確認是否有 PassengerId 重複的情況
print("重複的 PassengerId 數量：", df['PassengerId'].duplicated().sum())

# 檢查 Survived 欄位是否只包含 0 或 1
print("Survived 欄位的唯一值：", df['Survived'].unique())
```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 2 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  418 non-null    int64
 1   Survived     418 non-null    int64
dtypes: int64(2)
memory usage: 6.7 KB
None
重複的 PassengerId 數量： 0
Survived 欄位的唯一值： [0 1]
PS C:\Users\jing5\Documents\Hw2>
```

確認無誤將 IT_submission.csv 上傳到 kaggle

最後得到 0.7799 的 Public Score



[smo612/HW2-Logistic-regression]