題目：隨機森林分類預測模型

隨機森林是經典的監督式學習算法之一，請透過該算法進行 titanic_data.csv資料之分類預測。

- 相關套件提示
  - Python 套件：pandas, `from sklearn.ensemble import RandomForestClassifier`, `from sklearn.preprocessing import LabelEncoder`

10% 小題一：讀取資料

題目說明

- 請讀取 data資料夾內titanic_data.csv 資料集

答案示意

- 前五筆資料的輸出應如下(示意圖並非與答案完全一致)：

	pclass	survived	sex	age	sibsp	parch	fare	embarked
0	Upper	Yes	female	39	0	0	79	S
1	Upper	Yes	male	8	1	2	58	S
2	Upper	No	female	23	1	2	58	S
3	Upper	No	male	41	1	2	58	S
4	Upper	No	female	33	1	2	58	S

20% 小題二：切分訓練集與測試集

題目說明

- 請以資料筆數的 70%和 30%將 iris 資料隨機抽樣分為兩份，訓練集與測試集
  - 因具有隨機性質，抽樣結果可能不同，這部分不硬性規定，但若希望與答案示意結果相同可參考下方設定：
    - Python: `random_state=1` (此部分非必要執行)

答案示意

- (Python) 輸出應如下：

```
## 訓練集

##      pclass survived     sex  age  sibsp  parch  fare embarked
## 35    Upper      Yes  female   27      0      1   163        S
## 875   Lower       No    male   20      0      0   213        S
## 604   Lower      Yes  female   69      2      1    69        C
## 133   Upper       No    male   57      0      0     2        S
## 281   Upper      Yes  female   43      0      0    35        C
## ..      ...      ...     ...  ...    ...    ...   ...      ...
## 138   Upper       No    male   60      0      0   148        S
## 51    Upper      Yes  female   15      1      2    25        S
## 624   Lower       No    male   57      1      1    51        Q
## 275   Upper       No    male   70      1      1    80        C
## 503  Middle       No    male   33      0      0    26        S
##
## [732 rows x 8 columns]
```

```
## 測試集

##      pclass survived     sex  age  sibsp  parch  fare embarked
## 10    Upper       No    male   66      1      0    85        C
## 15    Upper       No    male   31      0      1    91        C
## 20    Upper      Yes  female   66      1      1   161        S
## 21    Upper      Yes    male   34      0      0   118        C
## 22    Upper      Yes  female   60      0      0    85        C
## ...     ...      ...     ...  ...    ...    ...   ...      ...
## 1034  Lower      Yes  female   66      1      0   189        S
## 1035  Lower       No    male   38      0      0    61        S
## 1037  Lower       No    male   36      0      0   253        S
## 1043  Lower       No    male   35      0      0   195        C
## 1045  Lower       No    male   39      0      0   223        S
##
## [314 rows x 8 columns]
```

30% 小題三：模型配適

題目說明

- 使用訓練集進行隨機森林模型配適
  - 自變數(x)：除了 survived 欄位，皆為自變數
  - 應變數(y)：survived 欄位

答案示意

- (Python) 配適後模型的輸出應如下：
  - 僅僅示意於 Python 中直接將模型輸出的結果，主要會看過程是否正確

```
## RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
##                        max_depth=None, max_features='auto', max_leaf_nodes=None,
##                        min_impurity_decrease=0.0, min_impurity_split=None,
##                        min_samples_leaf=1, min_samples_split=2,
##                        min_weight_fraction_leaf=0.0, n_estimators=10,
##                        n_jobs=None, oob_score=False, random_state=None,
##                        verbose=0, warm_start=False)
```

20% 小題四：預測

題目說明

- 以配適後模型進行測試集資料之預測
  - 放入模型的應為測試集資料的 speed 欄位，並得到相同數量的預測結果

答案示意

- (Python) 預測結果應如下：
  - 因為抽樣與訓練過程是隨機，**數值僅供參考**，主要會看過程是否正確

```
## array([0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1,
##        0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1,
##        1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0,
##        0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1,
```

20% 小題五：評估

題目說明

- 為評估預測結果與實際結果之誤差，請計算其準確度(accuracy)。

答案示意

- (Python) 準確度如下：
  - 因為抽樣與訓練過程是隨機，**數值僅供參考**，主要會看過程是否正確

```
## 0.732484076433121
```