

# 清大工工系 職涯導航儀

< 幫你找出屬於你的未來  >

第5組

109034036 柯妤萱

109034069 方弘德

110034539 廖滄舜





# Contents



- 背景介紹與問題分析
- 方法介紹與模型訓練績效
- Web
- 結論



01

# 背景介紹與 問題分析

# 背景介紹

## ■ 背景介紹:

- ✓ 剛踏入社會的新鮮人仍不清楚有哪些是適合自己的工作，十分迷惘
- ✓ 留言處有許多不真實的意見、或是「暗黑」他人的狀況發生
- ✓ 藉由使用機器學習模型、幫助使用者可以根據自己的能力、特質去找尋最適合自己的工作。



Dcard

<https://www.dcard.tw> > Dcard > 研究所

### [工工/工管出路- 研究所板](#)

2023年7月24日 — 工工相關的話，Planner、IE、智慧製造、生管品管物管都是自己的Domain，其他工科相關的EE、PE、機構，如果自己學校的工工有工院相關必修也可以走。最後是 ...



Dcard

<https://www.dcard.tw> > Dcard > 科技業

### [工業工程出路- 科技業板](#)

2023年10月14日 — 大家好~ 第一次發文如果有違反板規再麻煩提醒！，目前在讀四大工工碩（大學也是四大工工系的）預計會準時畢業，因為這兩年科技業都蠻凍的， ...



Dcard

<https://www.dcard.tw> > Dcard > 科技業

### [工工所出路請益- 科技業板](#)

其實類似文章也有幾篇，不過有些似乎消失了，但小弟還是想問一下未來的職涯發展（各位前輩手下留情TT，小弟剛上研究所一直在思考之後要往哪個方向發展，目前看到的主要 ...



國立高雄科技大學 工業工程與管理系



16

我以為工管的取代性超級高，什麼都學、什麼都不精通

B2 · 2023年6月12日 14:41 回覆

— 查看其他 7 則留言



國立臺灣大學

乖乖蹲製造帶線ㄅ

B1 · 2023年10月15日 11:41 回覆

— 查看其他 1 則留言

# 問題分析

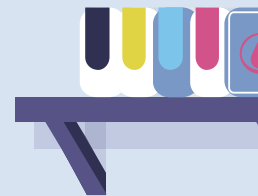


Who	<input type="checkbox"/> 清大工工系畢業學生
When	<input type="checkbox"/> 畢業、求職期間
Why	<input type="checkbox"/> 網路上眾說紛紜、且沒有一定的可信度 <input type="checkbox"/> 許多人畢業之後會開始陷入迷惘、找不到努力的目標 <input type="checkbox"/> 幫助使用者根據能力找到適合自己的工作
What	<input type="checkbox"/> 畢業出路推薦
Where	<input type="checkbox"/> 本網站的預測專區
How	<input type="checkbox"/> 網站結合 random forest 模型

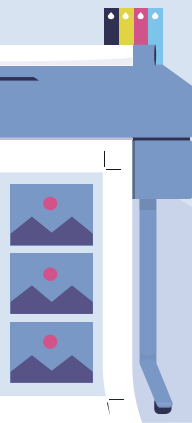


02

方法介紹

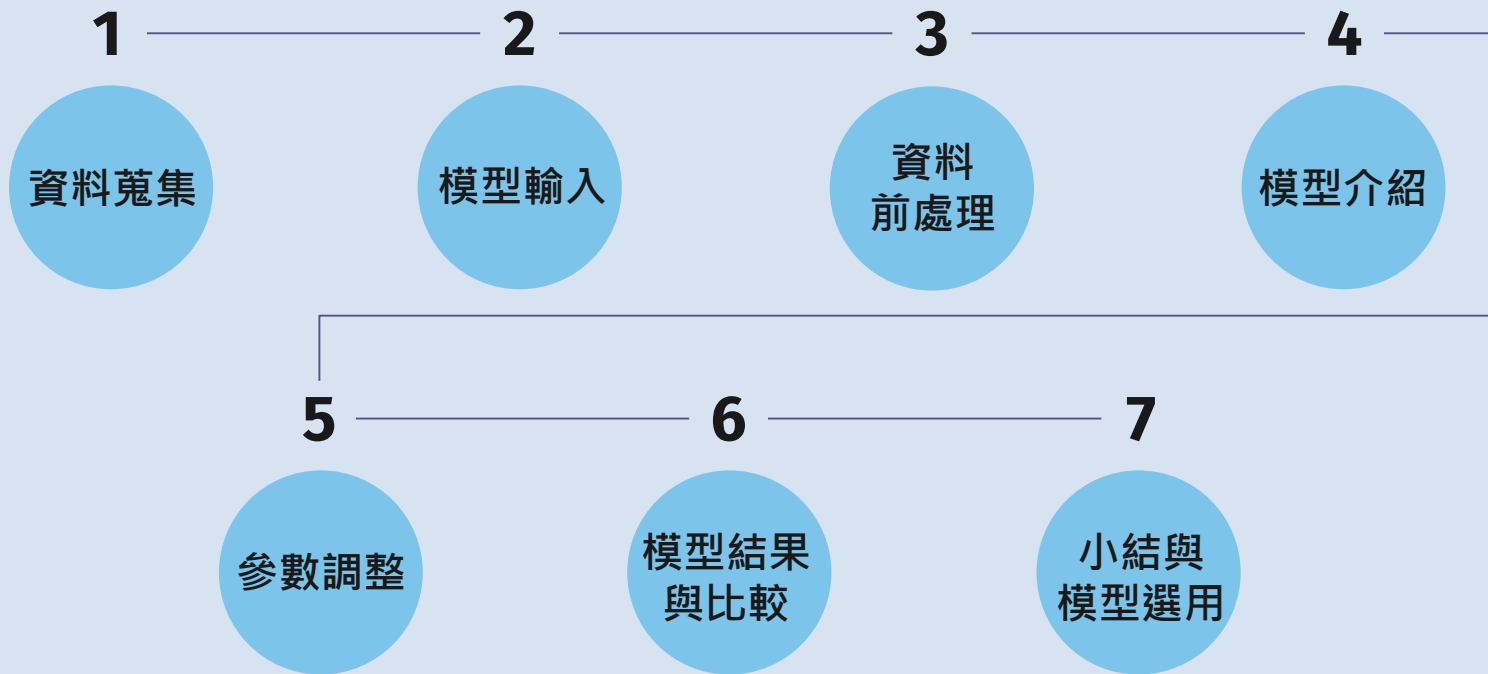


# 深度學習模型





# 模型要素







# 資料蒐集

- 本計畫與系辦職員進行合作，去除識別個人身分的特徵後獲得成績及任職資料，並與線上開源資料(Career-Prediction-System)結合，做為學生職業推薦系統。

Operations Research I	作業研究一的學期成績
Operations Research II	作業研究二的學期成績
Engineering Economics	工程經濟的學期成績
Quality Management	品質管理的學期成績
Production Control	生產計劃與管制的學期成績
Statistics	工程統計的學期成績
Human Factors Engineering	人因工程一的學期成績
Programming	計算機程式語言的學期成績
Job Role	此模型推薦的工作類別

名稱	說明
Logical quotient rating	邏輯能力
coding skills rating	程式能力
public speaking points	公開演講能力
self-learning capability?	自學能力
reading and writing skills	閱讀及寫作能力
memory capability score	記憶能力
Smart Ability score	聰明指數
Technical Skill Score	使用新技術的能力
Hard/Smart worker	努力型工作者或是 聰明型工作者





# 資料集介紹

- 資料筆數：482筆
- 工作類別：6種

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>1. 人因相關：34筆</li> <li>2. 智慧製造相關：204筆</li> <li>3. 產品相關：54筆</li> </ul> | <ul style="list-style-type: none"> <li>4. 資料科學相關：51筆</li> <li>5. 最佳化相關：67筆</li> <li>6. 品管相關：73筆</li> </ul> |
|--|--|

Logical coding	public	self-learning	reading	memory	Smart	Ab	Technical	Hard/Smart worker	Suggested Job Role	Operator	Operator	Engineer	Quality	Producti	Statisti	Human Fac	Program
8	2	10	5	9	8	6	5	Hard Worker	Human Factors Engineer	61	68	65	86	97	69	94	67
7	5	10	2	8	9	6	5	Hard Worker	Human Factors Engineer	67	64	70	82	92	60	90	66
4	6	9	8	6	6	8	0	Smart Worker	Human Factors Engineer	63	60	75	70	80	89	99	74
3	2	9	7	6	6	2	3	Smart Worker	Human Factors Engineer	71	71	98	91	97	67	93	66
8	1	10	4	6	8	2	5	Smart Worker	Human Factors Engineer	78	72	79	75	99	87	97	74
4	4	7	4	7	9	4	3	Smart Worker	Human Factors Engineer	77	60	76	82	94	88	93	60
4	2	7	2	6	9	4	3	Hard Worker	Human Factors Engineer	64	63	77	87	95	93	97	75
8	6	10	2	8	7	6	2	Hard Worker	Human Factors Engineer	75	78	61	91	93	95	98	62
7	1	8	3	8	9	5	1	Smart Worker	Human Factors Engineer	77	60	71	74	85	71	94	63
4	1	10	3	6	7	7	0	Smart Worker	Human Factors Engineer	75	68	64	87	94	85	92	79
3	4	8	5	6	7	3	3	Hard Worker	Human Factors Engineer	64	67	81	72	100	94	95	76
5	1	9	3	9	9	2	1	Hard Worker	Human Factors Engineer	72	68	80	99	95	63	100	64
5	3	6	7	7	7	4	2	Hard Worker	Human Factors Engineer	64	78	100	78	82	89	99	60

# 模型輸入



名稱	說明	數值範圍
Logical quotient rating	邏輯能力	[1,10]
coding skills rating	程式能力	[1,10]
public speaking points	公開演講能力	[1,10]
self-learning capability?	自學能力	[1,10]
reading and writing skills	閱讀及寫作能力	[1,10]
memory capability score	記憶能力	[1,10]
Smart Ability score	聰明指數	[1,10]
Technical Skill Score	使用新技術的能力	[1,10]
Hard/Smart worker	努力型工作者或是 聰明型工作者	Hard worker/ Smart worker (One-hot encoding)



# 模型輸入



Operations Research I	作業研究一的學期成績	[0,100]
Operations Research II	作業研究二的學期成績	[0,100]
Engineering Economics	工程經濟的學期成績	[0,100]
Quality Management	品質管理的學期成績	[0,100]
Production Control	生產計劃與管制的學期成績	[0,100]
Statistics	工程統計的學期成績	[0,100]
Human Factors Engineering	人因工程一的學期成績	[0,100]
Programming	計算機程式語言的學期成績	[0,100]
Job Role	此模型推薦的工作類別	共有六種職業類別



# 資料前處理

- **特徵選擇**：透過組員間腦力激盪，從系定必修擷取核心課程  
如：作業研究、生管、品管等
- **特徵縮放**：由於原先成績資料為等級制，會將成績轉換為百分比制，如：A → 85
- **資料擴增**：透過增加雜訊及resampling方式擴增資料

Classes	Human factor	Smart manufacturing	Product manager	Data Scientist	Optimization	Quality control
Raw data	34	204	54	51	67	73
Resampling	204	204	204	204	204	204



# 資料前處理-資料擴增

## □ 增加雜訊

```
def add_noise(value, range_min, range_max):  
    if range_max > 10:  
        noise = np.random.normal(0, 3)  
        range_max = 100  
        range_min = 0  
  
    else:  
        noise = np.random.normal(0, 1)  
        range_max = 10  
        range_min = 0  
  
    noisy_value = value + noise  
    return max(min(noisy_value, range_max), range_min)
```



# 資料前處理-資料擴增



## □ Resampling

```
from sklearn.utils import resample

df = pd.read_csv('new_data.csv')

print(df['Suggested Job Role'].value_counts())

max_samples = df['Suggested Job Role'].value_counts().max()

df_human_factors = df[df['Suggested Job Role'] == 'Human Factors Engineer']
df_smart_manufacturing = df[df['Suggested Job Role'] == 'Smart Manufacturing Engineer']
df_product_manager = df[df['Suggested Job Role'] == 'Product Manager']
df_data_scientist = df[df['Suggested Job Role'] == 'Data Scientist']
df_optimization_engineer = df[df['Suggested Job Role'] == 'Optimization Engineer']
df_quality_control_engineer = df[df['Suggested Job Role'] == 'Quality Control Engineer']

df_human_factors_upsampled = resample(df_human_factors, replace=True, n_samples=max_samples, random_state=123)
df_smart_manufacturing_upsampled = resample(df_smart_manufacturing, replace=True, n_samples=max_samples, random_state=123)
df_product_manager_upsampled = resample(df_product_manager, replace=True, n_samples=max_samples, random_state=123)
df_data_scientist_upsampled = resample(df_data_scientist, replace=True, n_samples=max_samples, random_state=123)
df_optimization_engineer_upsampled = resample(df_optimization_engineer, replace=True, n_samples=max_samples, random_state=123)
df_quality_control_engineer_upsampled = resample(df_quality_control_engineer, replace=True, n_samples=max_samples, random_state=123)

df_balanced = pd.concat([df_human_factors_upsampled, df_smart_manufacturing_upsampled, df_product_manager_upsampled,
                        df_data_scientist_upsampled, df_optimization_engineer_upsampled, df_quality_control_engineer_upsampled])

balanced_file_path = 'balanced_data.csv'
df_balanced.to_csv(balanced_file_path, index=False)

print(f"Balanced data has been saved to {balanced_file_path}")
```



# 模型介紹



- 基於決策樹的集成學習方法
- 抗過擬合能力強
- 對缺失數據不敏感

- 有較好的泛化能力
- 可解釋性強
- 適合中小規模且高維的分類任務

隨機  
森林

MLP

SVM

KNN

- 能捕捉複雜的非線性關係
- 靈活性高，可以調整網路結構
- 適合非線性且複雜的分類任務

- 基於距離的實例學習算法
- 無需訓練過程
- 適合數據量較小且特徵較少的分類任務





# 模型介紹-隨機森林

- **集成學習方法**：Random Forest 是一種基於多個決策樹的集成學習方法。每棵樹由一個隨機選擇的特徵子集訓練而成，最終的預測結果是這些樹的預測結果的平均（回歸問題）或多數表決（分類問題）。
- **袋裝法（Bagging）**：使用袋裝法對數據進行多次重複抽樣，從而生成多個訓練數據集。
- **適用本問題**：
  1. 學生成績數據包含多個科目分數，Random Forest 可以有效地處理這種多特徵數據
  2. 可以提供每個特徵的重要性，幫助我們理解哪些科目對職涯規劃建議最重要





# 模型介紹-多層感知器

- **神經網路結構**：由輸入層、隱藏層和輸出層組成。隱藏層數量和每層的神經元數可以調整，以適應不同的問題。
- **反向傳播算法**：通過反向傳播算法（Backpropagation）訓練網路，調整權重以最小化損失函數。
- **適用本問題**：
  1. 學生成績和職涯規劃之間可能存在複雜的非線性關係，MLP 能有效捕捉這些關係
  2. 可以通過調整隱藏層和神經元數量來適應不同的數據特徵和問題需求





# 模型介紹-支援向量機

- SVM 通過找到高維空間中的最佳超平面來最大化類別間距離
- 使用核函數（如線性核、多項式核、RBF核等）來處理非線性分類問題
  
- 適用本問題：
  1. 能夠處理多個特徵的學生成績數據
  2. 在小樣本數據下也能有良好的表現
  3. 可以清楚地解釋分類邊界





# 模型介紹-K近鄰演算法

- 為一種基於距離的實例學習算法
- 將新數據點分類到與其最近的K個鄰居的多數類別中
  
- **適用本問題：**
  1. 算法易於理解和實現
  2. 無需訓練過程：直接利用訓練數據進行分類
  3. 適合於特徵較少的數據：適合於特徵較少的數據：學生成績的核心課程數據相對簡單





# 超參數說明-隨機森林

## □ Random forest

參數名稱	描述
n_estimators	森林中樹的數量
max_depth	每棵樹的最大深度
min_samples_split	每個節點至少需要的樣本數，以便進行分裂
min_samples_leaf	每個葉子節點最少的樣本數





# 超參數說明-多層感知器 & K近鄰演算法

## □ Multilayer perceptron

參數名稱	描述
hidden_layer_sizes	每個隱藏層中的神經元數
learning_rate_init	初始學習率

## □ K-nearest neighbors

參數名稱	描述
n_neighbors	K值，即參與分類的最近鄰居數量





# 超參數說明-支援向量機

## □ Support vector machine

參數名稱	描述
C	正則化 ( regularization ) 參數
kernel	核函數類型
degree	多項式核函數的度數
gamma	核函數的係數





# 超參數設定-default setting

參數名稱	超參數設定
n_estimators	100
max_depth	None
min_samples_split	2
min_samples_leaf	1
hidden_layer_sizes	(100,)

參數名稱	最佳化範圍
learning_rate_init	0.001
C	1.0
kernel	Linear
degree	3
gamma	scale
n_neighbors	5







# 超參數最佳化-Optuna

- ❑ 本計畫使用Optuna進行超參數最佳化。
- ❑ Optuna 為基於貝氏最佳化 ( Bayesian optimization ) 的套件，通常應用在機器學習中，用來找出最佳的超參數組合，比起使用網格搜索、隨機搜索的方式找，貝氏最佳化能更聰明的選擇超參數，可以從較少的試驗次數下找到更好的結果。

參數名稱	最佳化範圍
n_estimators	[10, 200]
max_depth	[2, 32]
min_samples_split	[2, 20]
min_samples_leaf	[1, 20]
hidden_layer_sizes	(50,), (100,), (50,50), (100,50), (100,100)

參數名稱	最佳化範圍
learning_rate_init	[1e-5, 1e-1]
C	[1e-5, 1e2]
kernel	'linear'、'poly'、'rbf'、'sigmoid'
degree	[2, 5]若kernel='poly'，則為3
gamma	'scale'、'auto'
n_neighbors	[1, 20]



# 超參數最佳化-Optuna

## □ Optuna超參數程式，以MLP為例

```
def mlp_objective(trial):
    hidden_layer_sizes = trial.suggest_categorical('hidden_layer_sizes', [(50,), (100,), (50,50), (100,50), (100,100)])
    learning_rate_init = trial.suggest_loguniform('learning_rate_init', 1e-5, 1e-1)
    mlp_classifier = MLPClassifier(hidden_layer_sizes=hidden_layer_sizes, max_iter=500, learning_rate_init=learning_rate_init, random_state=42)
    mlp_cv_scores = cross_val_score(mlp_classifier, X_train, y_train, cv=5)
    return mlp_cv_scores.mean()

# 執行 MLP 的超參數調整
mlp_study = optuna.create_study(direction='maximize')
mlp_study.optimize(mlp_objective, n_trials=50)

# 輸出最佳參數
print("Best parameters for MLP:", mlp_study.best_params)

# 使用最佳參數進行訓練和預測
best_mlp_classifier = MLPClassifier(hidden_layer_sizes=mlp_study.best_params['hidden_layer_sizes'],
                                    max_iter=500,
                                    learning_rate_init=mlp_study.best_params['learning_rate_init'],
                                    random_state=42)

best_mlp_classifier.fit(X_train, y_train)
y_pred_best_mlp = best_mlp_classifier.predict(X_test)
best_mlp_accuracy = accuracy_score(y_test, y_pred_best_mlp)
best_mlp_classification_report = classification_report(y_test, y_pred_best_mlp)
```



# 模型比較

- 本計畫將原始資料（未平衡資料前），並使用基礎模型參數（未用Optuna最佳化）與平衡資料及使用最佳參數組合比較。

	Random forest			MLP			SVM			KNN		
	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score
Unbalanced data with default parameter	0.75	0.77	0.75	0.76	0.76	0.75	0.74	0.76	0.74	0.75	0.74	0.74
Unbalanced data with optimizing parameter	0.7	0.76	0.73	0.85	0.82	0.8	0.71	0.78	0.74	0.84	0.8	0.77
Balanced data with default parameter	0.98	0.98	0.98	0.99	0.99	0.99	0.83	0.82	0.82	0.85	0.85	0.85
Balanced data with optimizing parameter	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97

# 模型比較



## □ Cross-validation accuracy

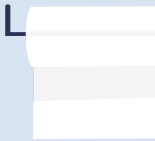
	Random forest	MLP	SVM	KNN
Unbalanced data with default parameter	<b>0.733</b>	0.7185	0.708	0.4783
Unbalanced data with optimizing parameter	<b>0.737</b>	0.6087	0.4224	0.5424
Balanced data with default parameter	<b>0.9804</b>	0.7908	0.8554	0.7255
Balanced data with optimizing parameter	<b>0.9779</b>	0.7917	0.848	0.9436



# 小結與模型選用

- ❑ 在數據平衡 ( resampling ) 過後，所有模型的Cross-validation accuracy皆有上升
- ❑ Random forest在不平衡資料及resampling過後，皆表現優於其他模型
- ❑ 根據實驗結果，本計畫將採用Random forest作為學生職業推薦之預測模型，並將使用預設之參數



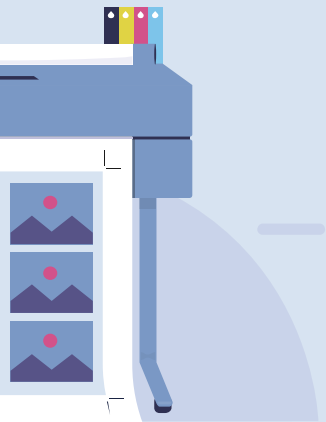


# 03

## Web



# Web-DEMO





# URL



1. myphpadmin 資料庫

[https://140.114.54.94/phpmyadmin/sql.php?db=2024iie\\_group5&table=jobstudent&pos=0](https://140.114.54.94/phpmyadmin/sql.php?db=2024iie_group5&table=jobstudent&pos=0)

2. 前端

[http://140.114.54.94/IIE\\_2024/group5/Project2front/JobAdvice.html](http://140.114.54.94/IIE_2024/group5/Project2front/JobAdvice.html)

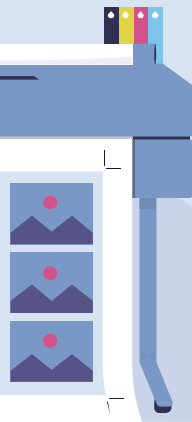
3. 後端

[http://140.114.54.94/IIE\\_2024/group5/project2back/index.html](http://140.114.54.94/IIE_2024/group5/project2back/index.html)



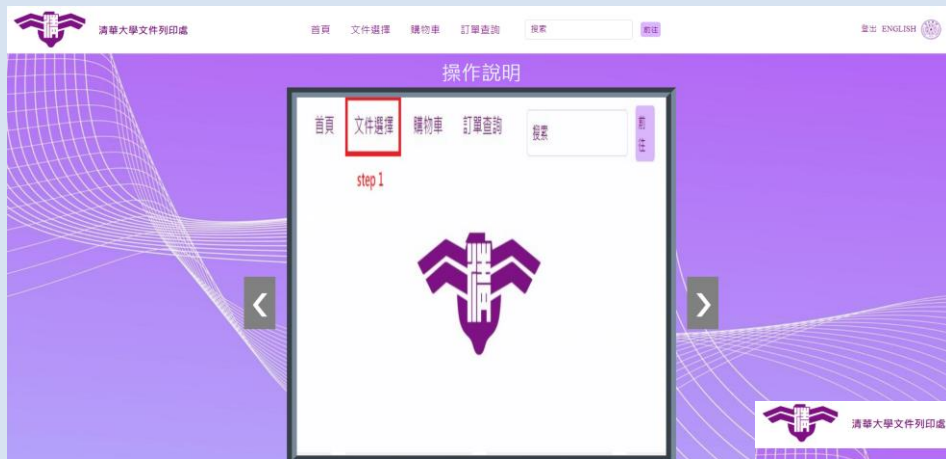


# Web-前端





# 網站功能介紹-首頁





# 網站功能介紹-選擇文件

學士班中文名次證明

\$20

放入購物車

學士班英文名次證明

\$20

放入購物車

學士班中文在學證明

\$20

放入購物車

學士班英文在學證明

\$20

放入購物車

學士班當學期(年)中文成績單

\$20

放入購物車

學士班當學期(年)中文成績單(含T分數)

\$15

放入購物車

學士班當學期(年)中文成績單(含修課相對成績)

\$15

放入購物車

學士班學生證遺失補發

\$200

放入購物車

學士班歷年中文成績單

\$20

放入購物車

學士班歷年中文成績單(含T分數)

\$20

放入購物車

學士班歷年中文成績單(含名次)

\$20

放入購物車

學士班歷年中文成績單(含名次·T分數)

\$20

放入購物車

# 網站功能介紹-購物車



清華大學文件列印處

首頁 文件選擇 購物車 訂單查詢  [前往](#) [登出](#) [ENGLISH](#)

產品型錄

產品	定價	數量	小計	變更數量
Chinese Grade Report	\$20	<input type="text" value="3"/>	\$60	<input type="button" value="修改"/>
Chinese Student Identification	\$20	<input type="text" value="1"/>	\$20	<input type="button" value="修改"/>
English Grade Report	\$20	<input type="text" value="1"/>	\$20	<input type="button" value="修改"/>

總金額 = 100

# 網站功能介紹-訂購資料填寫



清華大學文件列印處

首頁 文件選擇 購物車 訂單查詢  [前往](#) [登出](#) [ENGLISH](#)

產品型錄

請填入下列資料 (標示「\*」欄位請務必填寫)

*收件人姓名：	<input type="text" value="王小明"/>
電話：	<input type="text" value="0912-345678"/>

[更多關於我們](#)

# 網站功能介紹-訂購完成畫面



要打印時請掃描以下QR code或輸入訂單ID



恭喜您已經成功下訂，您的訂單如下：(請勿按重新整理鍵)

[回到網站繼續購物回到購物車](#)

訂單序號	產品	定價	數量	小計
20240418-898	Chinese Grade Report	\$20	3	\$60
20240418-898	Chinese Student Identification	\$20	1	\$20
20240418-898	English Grade Report	\$20	1	\$20

總金額 = 100

# 網站功能介紹-訂單查詢

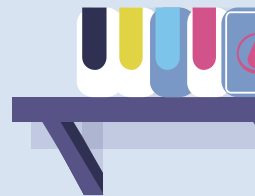


清華大學文件列印處首頁 文件選擇 購物車 訂單查詢  前往 登出 ENGLISH 

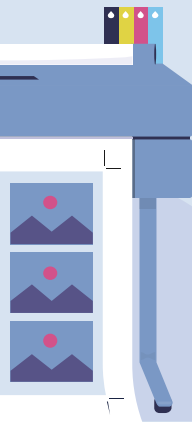


您的訂單如下：( 請勿按重新整理鈕 )

收件人	訂單序號	產品	定價	數量	下定日期	小計
ko	20240418-807	?????????	\$20	1	2024-04-18 23:04:08	\$20
						總金額 = 20



# Web-後端





# Web後端-訂單資料



清華大學文件列印處

首頁 訂單資料 用戶資料

搜索

前往

登出 ENGLISH



## 訂單修改

### 訂單

Show 10 entries

Search:

Order ID	Name	Phone	Document	Total Price	Quantity	Date
20240418-016	??	968	Chinese Grade Report	20	1	2024-04-18 09:30:49
20240418-016	??	968	Chinese Student Identification	20	1	2024-04-18 09:30:49
20240418-066	26	26	Chinese Grade Report	20	1	2024-04-18 09:57:11
20240418-083	jj	0	Chinese Student Identification	100	5	2024-04-18 01:59:05
20240418-097	jj	0	Chinese Student Identification	100	5	2024-04-18 01:58:05
20240418-107	rr	0	Chinese Grade Report	120	6	2024-04-18 00:06:07
20240418-108	?	968	Chinese Grade Report	40	2	2024-04-18 11:18:13
20240418-108	?	968	Chinese Student Identification	20	1	2024-04-18 11:18:13
20240418-116	jj	0	Chinese Student Identification	100	5	2024-04-18 01:46:34
20240418-129	jkkk	890	Chinese Student Identification	100	5	2024-04-18 01:17:44

Showing 1 to 10 of 60 entries

Previous 1 2 3 4 5 6 Next

### 取消訂單

\*訂單ID:  (注意：是訂單ID，不是賬號)

取消訂單

# Web後端-用戶資料



清華大學文件列印處

[首頁](#) [訂單資料](#) [用戶資料](#)

搜索

前往

登出 ENGLISH



## 清華列印系統用戶個人資料

所有儲存於資料庫之會員用戶資料皆出現於此

### 會員資料

Show 10 entries

Search:

Account	Password	Name	Email
109034002	1234	hsuan	hihi
109034003	1234	hsuan	hihi
109034004	1234	hsuan	hihi
109034005	1234	hsuan	hihi
109034006	1234	hsuan	hihi
109034007	1234	hsuan	hihi
109034008	1234	hsuan	hihi
109034009	0121	lmm	abc
109034036	ko	koo	koko
109340788	12	kay	kayyy

# Web後端-用戶資料



## 新增會員

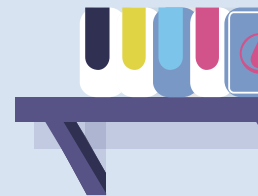
*帳號 :	<input type="text"/>	(請輸入完整學號)
*密碼 :	<input type="text"/>	(請使用英文或數字鍵)
*密碼確認 :	<input type="text"/>	(再輸入一次密碼)
*姓名 :	<input type="text"/>	
E-mail 帳號 :	<input type="text"/>	
<input type="button" value="加入會員"/> <input type="button" value="重新填寫"/>		

## 刪除會員

*帳號 :	<input type="text"/>
<input type="button" value="刪除會員"/>	

## 會員資料修改

*帳號 :	<input type="text"/>	
*密碼 :	<input type="text"/>	(請使用英文或數字鍵)
*密碼確認 :	<input type="text"/>	(再輸入一次密碼)
*姓名 :	<input type="text"/>	(請使用英文或中文)
*E-mail :	<input type="text"/>	
<input type="button" value="會員修改"/> <input type="button" value="重新填寫"/>		



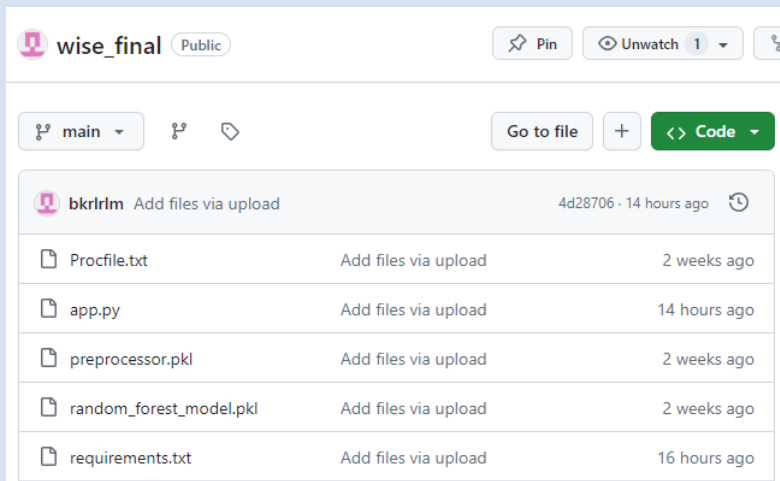
# Web— 職涯導航儀





# 網站與模型串接

- 使用Flask來處理預測的請求，再使用Github來進行版本控制、並將其應用部署到render上，此部分的主要架構有: preprocessor.pkl、random\_forest\_model.pkl、Procfile.txt、requirements.txt、app.py等部分。





# 1.preprocessor.pkl 與random\_forest\_model.pkl

- ❑ preprocessor.pkl將特徵進行標準化，使input給模型時可以保證是相同的transform方式。
- ❑ random\_forest\_model.pkl是主要模型部分，用來預測最佳推薦工作出路。

```
import joblib

joblib.dump(rf_classifier, 'random_forest_model.pkl')
joblib.dump(preprocessor, 'preprocessor.pkl')
```



## 2. Procfile.txt 與 requirements.txt

- ❑ Procfile.txt是用來指定應用入口點
- ❑ requirements.txt是為了列出執行此程式時，所需要的所有依賴 package，這些package會在部署時自動安裝。

```
Code Blame Raw Copy Download Edit Dropdown Code
```

```
1 web: gunicorn app:app
```

---

```
Code Blame Raw Copy Download Edit Dropdown Code
```

```
1 flask
2 flask-cors
3 joblib
4 pandas
5 scikit-learn==1.2.2
6 gunicorn
```

# 3.app.py

- ❑ 創建flask來處理網頁端的POST請求、以及利用json檔來做response。除此之外，我們也加入“CORS”來處理跨域的問題，讓此API可以被其他域訪問

```
1 from flask import Flask, request, jsonify
2 from flask_cors import CORS
3 import joblib
4 import pandas as pd
5 import logging
6
7 logging.basicConfig(level=logging.INFO)
8
9 try:
10     rf_classifier = joblib.load('random_forest_model.pkl')
11     preprocessor = joblib.load('preprocessor.pkl')
12 except Exception as e:
13     logging.error(f"Error loading model or preprocessor: {e}")
14     raise e
15
16 app = Flask(__name__)
17 CORS(app)
18
19 @app.route('/predict', methods=['POST'])
20 def predict():
21     try:
22         data = request.get_json()
23         logging.info(f"Received data: {data}")
24
25         required_fields = [
26             'Logical quotient rating', 'coding skills rating', 'public speaking points',
27             'self-learning capability', 'reading and writing skills', 'memory capability score',
28             'Smart Ability score', 'Technical Skill Score', 'Hard/Smart worker',
29             'Operations Research I', 'Operations Research II', 'Engineering Economics',
30             'Quality Management', 'Production Control', 'Statistics',
31             'Human Factors Engineering', 'Programming'
32         ]
33     ]
34
35     for field in required_fields:
36         if field not in data:
37             return jsonify({'error': f'{field} field is missing'}), 400
38
39     new_data_df = pd.DataFrame([data])
40     new_data_df = pd.get_dummies(new_data_df, columns=['Hard/Smart worker'], drop_first=True)
41
42     for col in preprocessor.feature_names_in_:
43         if col not in new_data_df.columns:
44             new_data_df[col] = 0
45
46     new_data_df = new_data_df[preprocessor.feature_names_in_]
47     new_data_scaled = preprocessor.transform(new_data_df)
48
49     prediction = rf_classifier.predict(new_data_scaled)
50     logging.info(f"Prediction: {prediction[0]}")
51
52     return jsonify({'predicted_job_role': prediction[0]})
53 except Exception as e:
54     logging.error(f"Error during prediction: {e}")
55     return jsonify({'error': str(e)}), 500
56
57 if __name__ == '__main__':
58     app.run(debug=True)
```







## 4.html/JS

- 利用html表單的方式蒐集使用者所輸入的值（對自身能力如公眾發言、程式能力、自我提升能力等等以及主科的成績進行評估與輸入。
- 利用 javascript 發送 Flask API，response 過後在網頁端顯示預測結果。

Programming 

Predicted Job Role 

不需填寫

[提交](#) [清空](#) [儲存系統](#)

若想將本次輸入與結果儲存到系統，請在點擊“提交”後，將工作職位複製貼到“Predicted Job Role”欄位，並點擊“儲存系統”

[更多關於我們](#)


- [首頁](#)
- [文件選擇](#)
- [如何選擇文件](#)
- [如何查詢歷史記錄](#)
- [若有疑問，請聯繫我們](#)
- [總機: 03-5715131](#)

## 4.html/JS

□ 結果返回之示意圖：




77

Programming 

90

Predicted Job Role 

不需填寫

Predicted Job Role: Data Scientist 



## 5. 選擇是否將成績與預測結果儲存到後臺系統

- 可選擇是否保存該次的輸入與預測結果到後台系統
- 與先前介紹之Web後端相同，數據資料將儲存到phpMyAdmin的資料庫



04

# Business model



# Business model



REVENUE STREAMS	CHANNELS	VALUE PROPOSITION	CUSTOMER RELATIONSHIPS	KEY ACTIVITIES
<p>廣告費用： 在我們的網站上，且在不影響用戶體驗下設置廣告區。</p> <p>訂閱模式： 使用者付費可以享有我們的免廣告服務。未來如果本網站有結合更多的應用（如：MBTI服務），可以唯一享有此新功能</p>	<p>網路： 使用者透過網際網路連線就可以使用我們的系統</p> <p>KEY RESOURCES</p> <p>PaaS 雲端服務： 本project是用render來部署網站</p> <p>實驗室的伺服器： 本網站就在此伺服器下開發</p>	<p>多元化網站： 我們的網站不僅提供輔助列印成績單的功能，還可以讓使用者可製成預測適合的職涯</p> <p>客製化預測： 系統會根據個人特質和主修成績提供量身訂做的職涯建議</p> <p>友善介面： 我們介面簡單易懂、方便操作</p>	<p>技術支援： 會提供聯絡方式來幫助使用者解決系統上的問題以及不便之處</p> <p>自助服務： 使用者可自行在我們開發的網站處理大部分需求</p> <p>COSTOMER SEGMENT</p> <p>學生： 需要申請相關文件還有對職涯迷惘的學生們</p>	<p>系統開發： 開發預測模型並整合到網站之中</p> <p>數據分析： 根據訓練資料做清洗以及前處理</p> <p>訓練模型： 需比對多種模型效能來決定最終模型</p>
COST STRUCTURE		KEY PARTNERS		
<p>運營成本： 如果使用者超過一定的數量，我們就需要使用付費版的render來協助我們網站的營運</p>		<p>清華大學工工系辦： 需要他們協助提供資料</p>		



04

結論

# 結論



✓ 清大工業工程與工程管理學系學生工作推薦服務:

- 前端網站結合機器學習模型

- 學生輸入主科成績與自身軟技能的評分，通過模型根據以上數據推薦工作職位

✓ 未來展望:

- 結合更多學生相關申請事務於我們的系統之中、為清大學生提供更多的便利

- 提升模型泛用性，讓全台灣的工業工程與工程管理學系學生透過我們的模型進行工作職位的推薦