

# 2024 第五屆T大使 數位之星 實作成果組

KPMG

安侯永續發展顧問(股)有限公司

---

隊名：早安侯塞雷

彭伊璪、謝承諭、王郁涵

陳柏臻、陳峻凱



# Agenda

- 01 數位轉型需求
- 02 解決方案
- 03 數位工具運用說明
- 04 質化 / 量化目標達成狀況
- 05 未來精進建議
- 06 團隊分工
- 07 與企業如何協作



國內大型金控

## 課題描述

- AI發展日新月異，議題與應用方式眾多，許多企業均期望可以了解AI應用與既有產業如何結合，整合進既有的業務流程當中。

## 實作內容

- 了解數位發展與政策發展適用哪些AI技術及如何應用，例如：AI金融國內外應用發展調研、AI工具研究與實際應用

## 目標

- 提供金融業AI發展策略建議

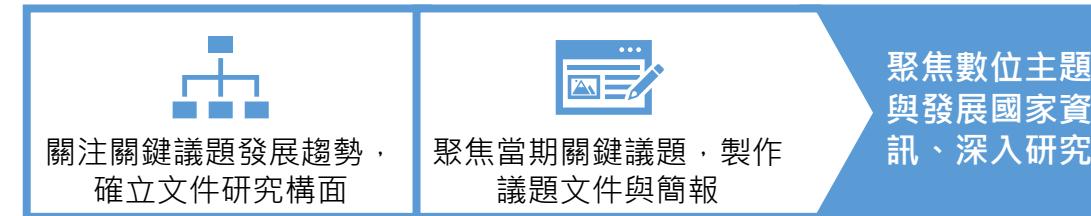
## 解決方案與預期成果

目前透過連結型企業方法論(Connected Enterprise)、Microsoft PPT進行國內外案例初步研究調查，了解金融機構的AI應用策略與洞察，後續將結合RAI等工具的實際應用，協助進行案例發想與研究。

### 1 廣泛的資訊調研 Broadly Scan



### 2 依發展現況 聚焦主題與深入研究



### 3 RAI工具應用實例



# 銀行決策支援系統 ( DSS )

## 數據整合



# 銀行決策支援系統 ( DSS )

- 收集客戶資料庫、交易記錄

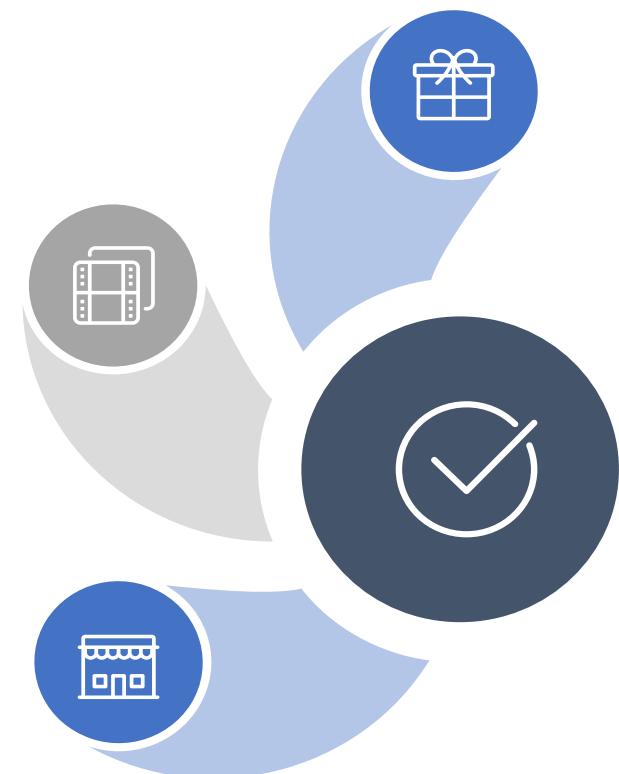
數據整合

- 機器學習：依據行業資料特性，選擇適合模型
- 客群分析：預測客戶行為
- 識別行為模式：以針對特定群體推廣產品。
- RAI工具：確保模型的決策過程可以被理解和追溯，且評估模型是否存在偏差。

分析工具

- 藉由分析工具，獲得更加詳細和精確的數據分析，可了解各種決策選項的潛在影響

決策支援



# 數位工具運用說明

KPMG  
課程

Python 實作

RAI工具



工具介紹



課程教學

**XGBoost**



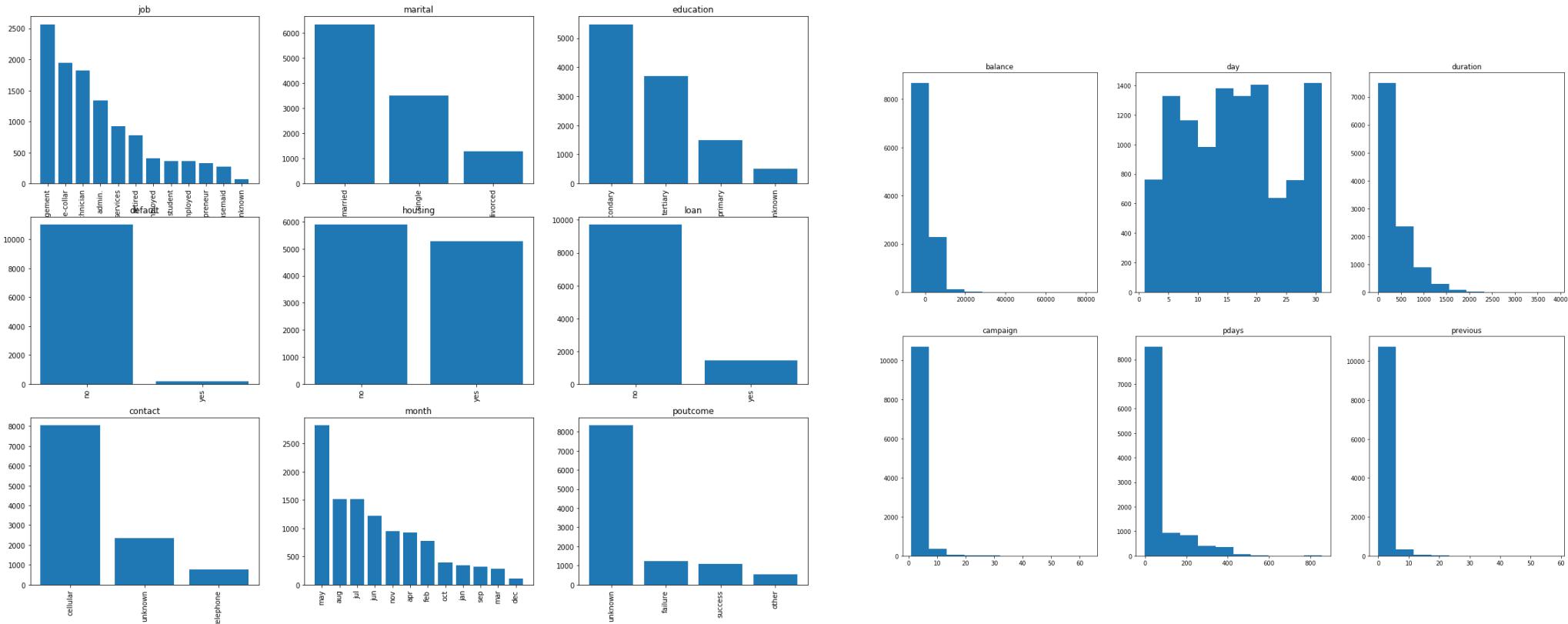


- 集中來自不同來源的數據，如客戶資料庫、交易記錄和外部市場數據。確保數據的一致性和準確性，為決策過程提供可靠的基礎。

# 數據資料整合/確認

- 使用 kaggle 上的 bank marketing dataset

age  
job  
marital  
education  
default  
balance  
housing  
loan  
contact  
day  
month  
duration  
campaign  
pdays  
previous  
poutcome  
deposit



數列資料

數值資料



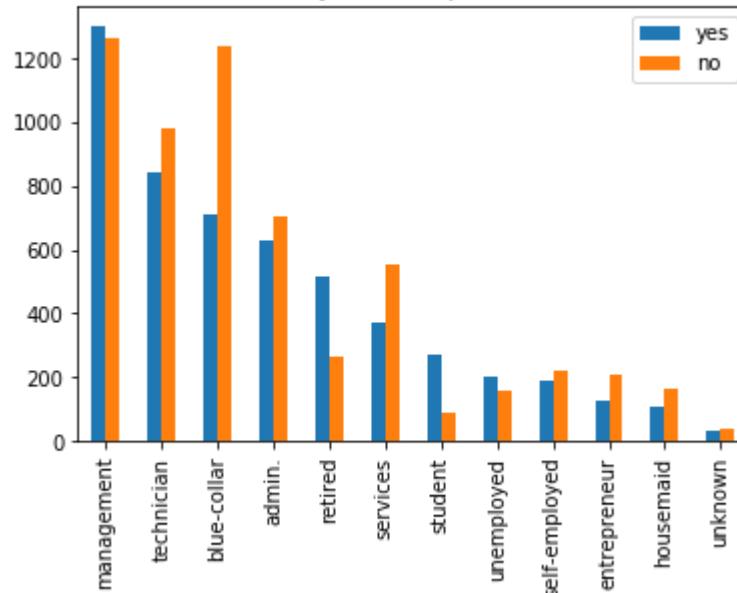
- 應用如群集分析和機器學習，對客戶數據進行細分，以識別不同的客戶群體和行為模式。

# Deposit

- 表示是否購買「定期存款或儲蓄存款」

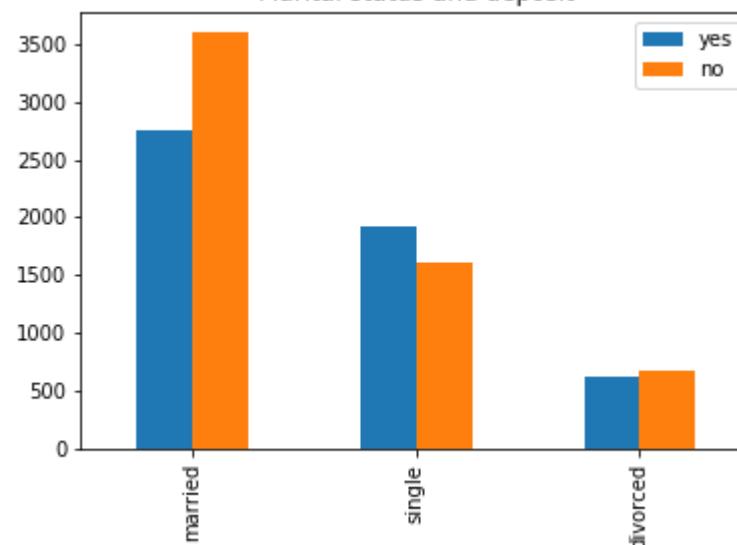
職業

Job and deposit



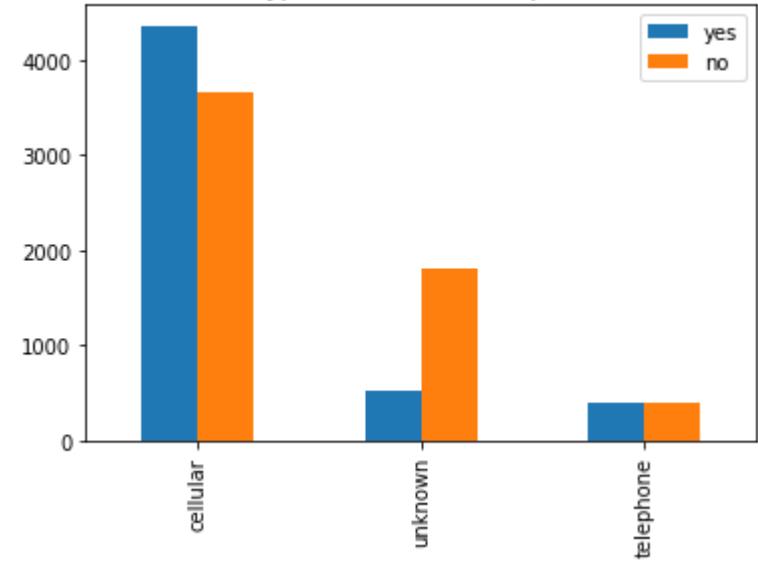
婚姻

Marital status and deposit



聯絡方式

Type of contact and deposit

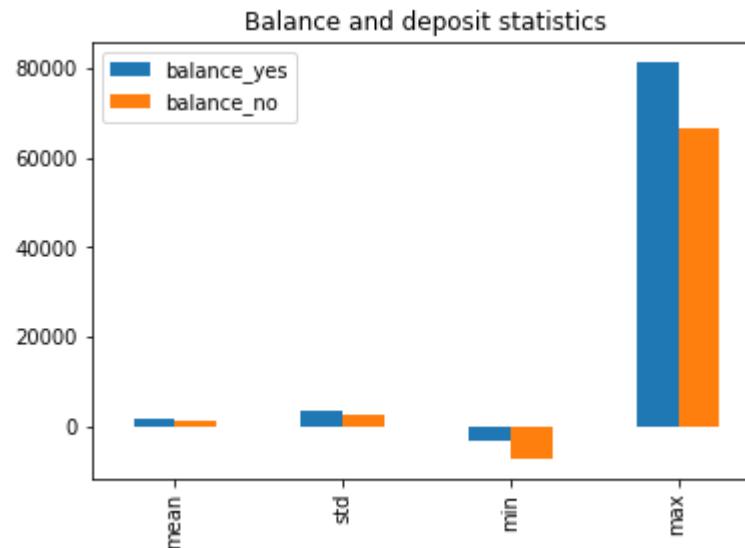


- 職業是「blue-collar」和「services」；婚姻狀態為「已婚」；聯絡方式是「cellular」者不太可能購買定存

## 數值資料

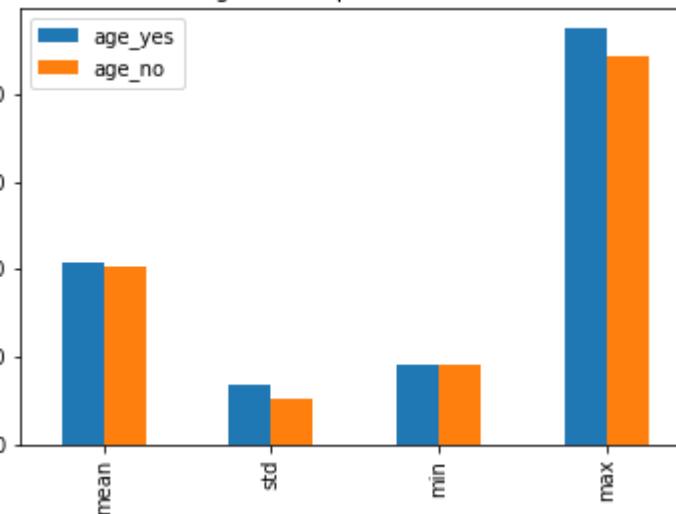
- 藉由觀察數字相關資料的特性，找出相對應的客群

### 存款

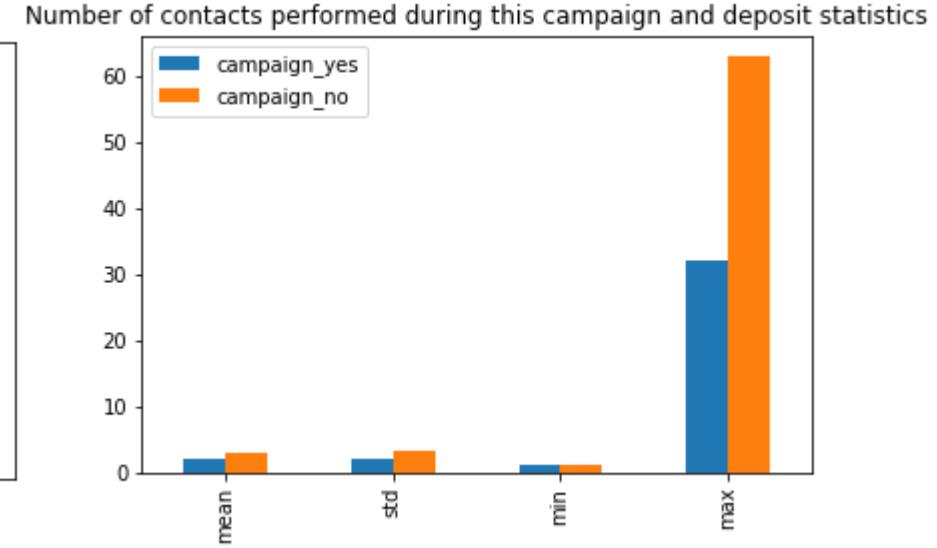


### 年齡

Age and deposit statistics



### 聯絡次數



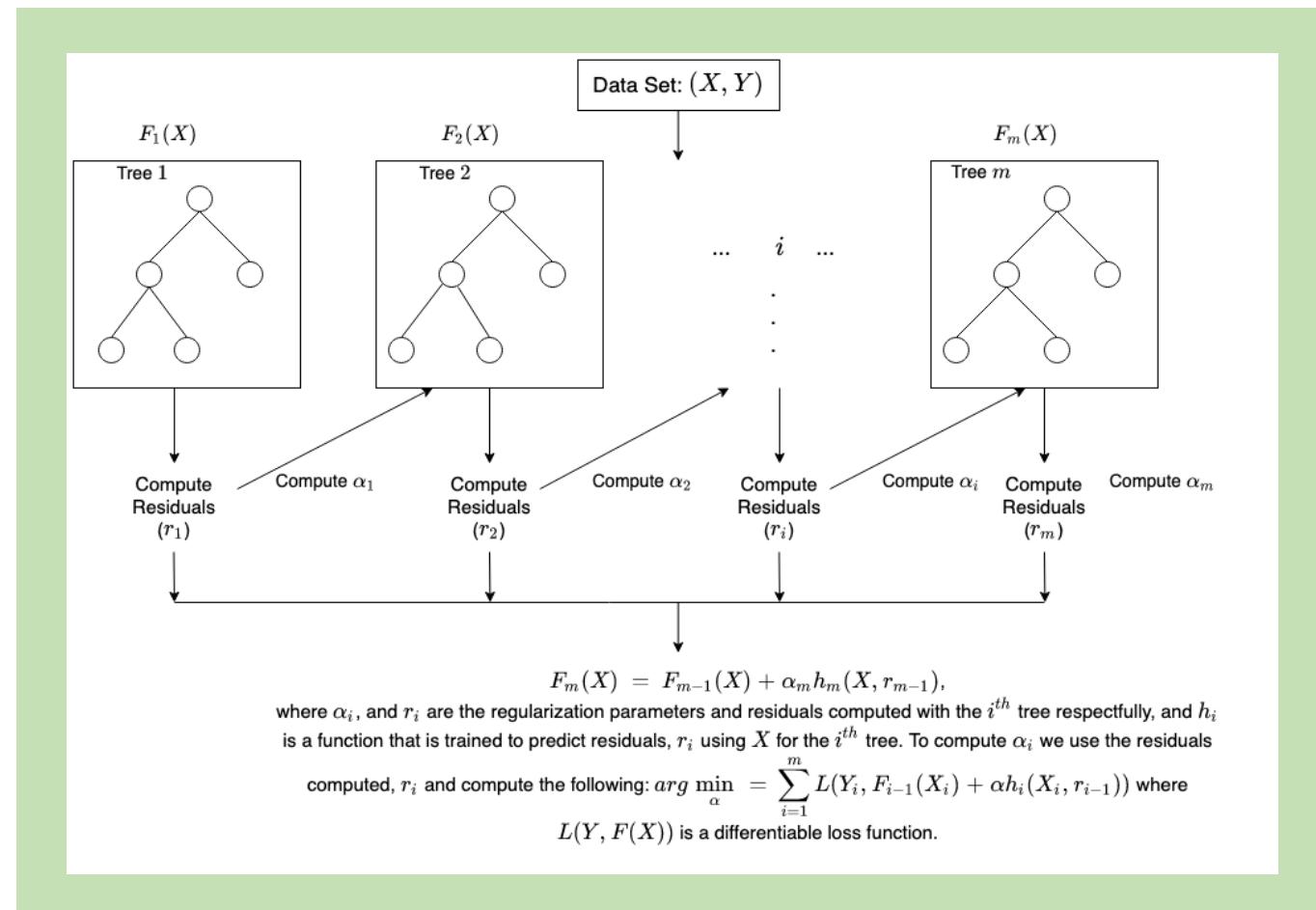
- 以下條件的人比較容易買定存：銀行存款較多、年齡越大
- 聯絡次數較多，拒絕購買機率高



- 以銀行決策支援系統（DSS）的概念出發，由市場銷售活動數據集訓練模型，制定下次活動的銷售策略。

# 模型選擇 - XGBOOST

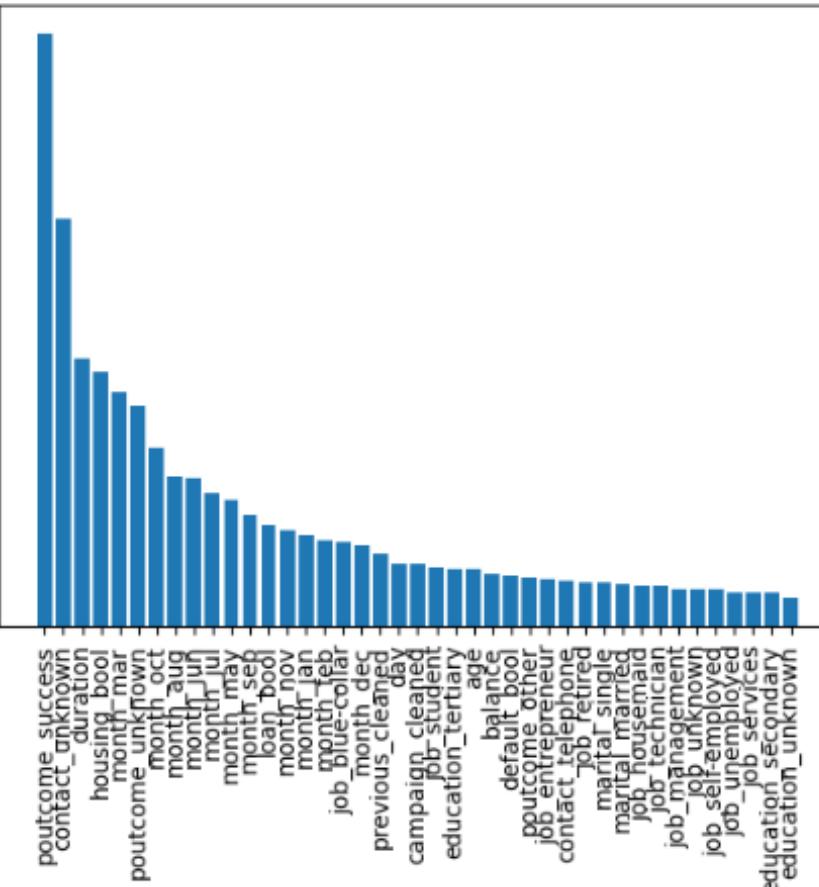
1. 效率和性能 - 基於 Gradient Boosting 的分布式樹模型演算法
2. 模型精確度 - 競賽表現優
3. 靈活性 - 支持多目標函數和多種評估準則
4. 缺失值 - 可以自動處理缺失
5. 正則化 - 包括L1、L2 正則化，避免過擬合
6. 可解釋性 - 特徵重要性評分



## 機器學習

- 藉由觀察數列資料特性，找出客群

Feature importances (XGB)



```
# 找特徵
headers = ["name", "score"]
values = sorted(zip(X_train.columns, xgb.feature_importances_), key=lambda x: x[1] * -1)
xgb_feature_importances = pd.DataFrame(values, columns = headers)

#plot feature importances
x_pos = np.arange(0, len(xgb_feature_importances))
plt.bar(x_pos, xgb_feature_importances['score'])
plt.xticks(x_pos, xgb_feature_importances['name'])
plt.xticks(rotation=90)
plt.title('Feature importances (XGB)')

plt.show()
```

- 最重要的特徵包括：客戶的帳戶餘額、年齡、在這次活動中聯繫次數和聯繫持續時間等。
- 建模的主題會是：年齡、存款、聯繫次數和是否購買定存的關係



- 根據不同商品性質，群體的購買行為、偏好和反應，用以制定更精準的市場策略和個性化的產品推廣，以提高銷售效率並吸引更多潛在客戶。

## 數據計算

- 經由計算，找出特定的年齡、財務狀況，及銷售時應注意的部分，以達成更好的銷售目標

```
#introduce new column 'age_buckets' to ''
df_new['age_buckets'] = pd.qcut(df_new['age'], 20, labels=False, duplicates = 'drop')

#group by 'balance_buckets' and find average campaign outcome per balance bucket
mean_age = df_new.groupby(['age_buckets'])['deposit_bool'].mean()

#plot
plt.plot(mean_age.index, mean_age.values)
plt.title('Mean % subscription depending on age')
plt.xlabel('age bucket')
plt.ylabel('% subscription')
plt.show()
```

```
df_new = cleaned_df.copy()

#introduce new column 'balance_buckets' to ''
df_new['balance_buckets'] = pd.qcut(df_new['balance'], 50, labels=False, duplicates = 'drop')

#group by 'balance_buckets' and find average campaign outcome per balance bucket
mean_deposit = df_new.groupby(['balance_buckets'])['deposit_bool'].mean()

#plot
plt.plot(mean_deposit.index, mean_deposit.values)
plt.title('Mean % subscription depending on account balance')
plt.xlabel('balance bucket')
plt.ylabel('% subscription')
plt.show()
```

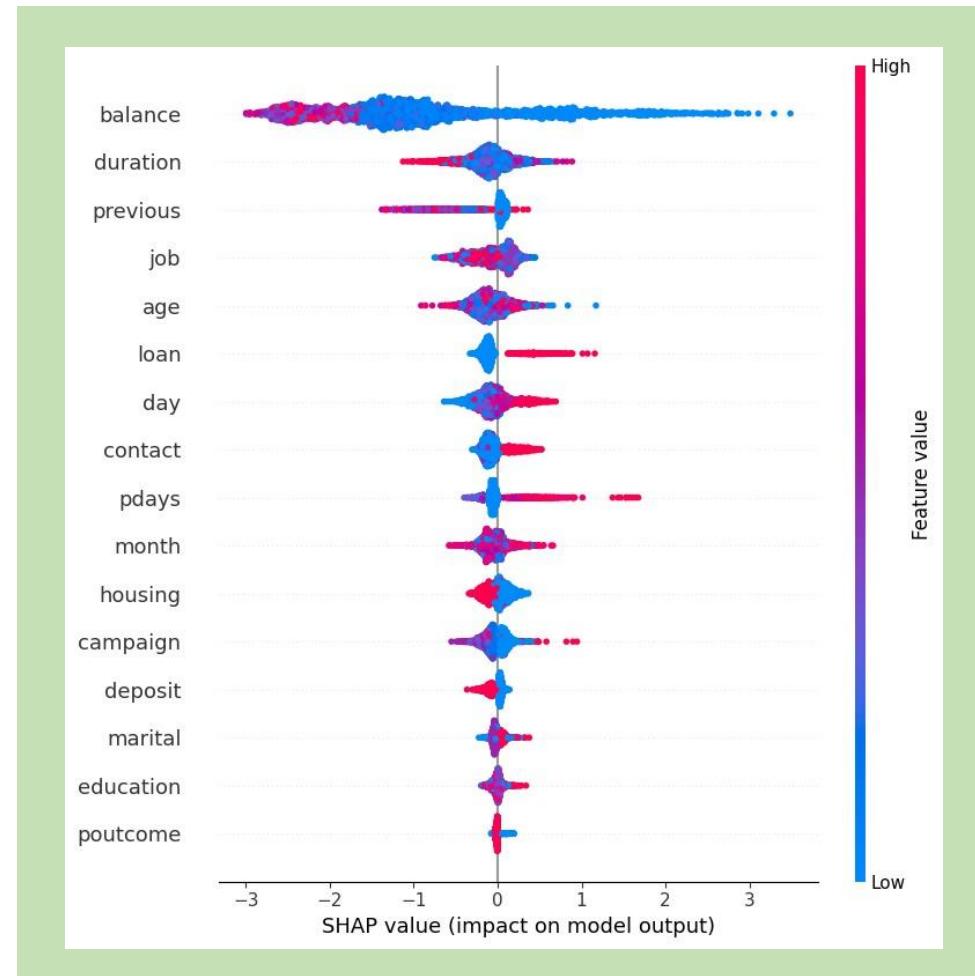
```
#introduce new column 'age_buckets' to ''
df_new['age_buckets'] = pd.qcut(df_new['age'], 20, labels=False, duplicates = 'drop')

#group by 'balance_buckets' and find average campaign outcome per balance bucket
mean_age = df_new.groupby(['age_buckets'])['deposit_bool'].mean()

#plot
plt.plot(mean_age.index, mean_age.values)
plt.title('Mean % subscription depending on age')
plt.xlabel('age bucket')
plt.ylabel('% subscription')
plt.show()
```

## SHAP可視化

- Balance  
    餘額越高，越傾向於正類別（購買定存）
- Duration  
    單次通話時間越長，越可能成功交易
- Previous  
    過去聯繫次數較多的樣本，會對模型有負面影響。
- job、marital、education：  
    這些特徵點的顏色深淺不同，表示不同群體在決策過程中的差異。



## 量化/質化

## 量化

## 質化

針對  
帳戶餘額超過**1490** 歐元、31 – 56 歲  
且  
活動期間與客戶聯繫次數**少於 4** 次  
可達**80%**以上的定存購買率

- 增強決策精確性
- 提高操作效率
- 客戶關係管理
- 銷售策略規劃

## 未來精進建議



- 強化模型
- 強化數據
- 數據更新
- 效益評估

## 團隊分工

**陳柏臻**

- 案例調查
- 資料蒐集
- 簡報製作

**王郁涵**

- 案例調查
- 資料蒐集
- 簡報製作

**彭伊璽**

- 資料收集
- 客戶分析
- 解釋模型

**謝承諭**

- 案例分析
- 建立模型
- 制定策略

**陳嶽凱**

- 案例分析
- 建立模型
- 制定策略

# 企業協作與溝通

## 企業方法論

數位轉型經驗分享，同時以課程讓我們熟悉企業運作、作業流程，充實顧問須具備的能力。

## RAI應用-工具學習

以金管會指引為出發，教授可解釋、公平性、可靠性、透明性的RAI工具。

## 國內外AI趨勢分析

調查國內外政府因應AI趨勢，為了人才招募所作的相應措施，並進行簡報教育訓練。



## RAI應用-案例實作

完整的金融業RAI應用專案  
帶領學員進行python實作

# THANK YOU

