

# **銀行客戶流失預測專案提供有效的業務策略與分析方法**



# 1 數據預處理與分析

# 針對銀行客戶流失的數據清洗與處理方法如何實施

## 數據前處理程序

利用Pandas和Numpy進行資料清洗，確保數據質量，提升預測準確性

## 類別變數編碼

Label Encoding技術處理類別變數，增強模型的解釋力和效能

## 訓練測試集分割

合理劃分訓練集與測試集，確保模型的泛化能力及評估的準確性

# 使用Pandas和Numpy進行流失預測的數據分析流程

## 選擇基準模型

Logistic Regression作為基準模型，擁有良好的解釋性能，便於理解

## 非線性模型應用

隨後應用Random Forest模型，處理非線性特徵及變數間的交互作用

## 模型表現評估

使用Accuracy、Precision、Recall和F1score評估模型的綜合表現與效果

## 結果視覺化方法

透過混淆矩陣和指標長條圖展示模型預測結果，幫助業務分析



# 如何有效進行類別變數的Label Encoding以提升模型性能

## 特徵重要性分析

識別影響預測的關鍵變數，包括信用評分、活躍會員標識和帳戶餘額

## 高風險客群識別

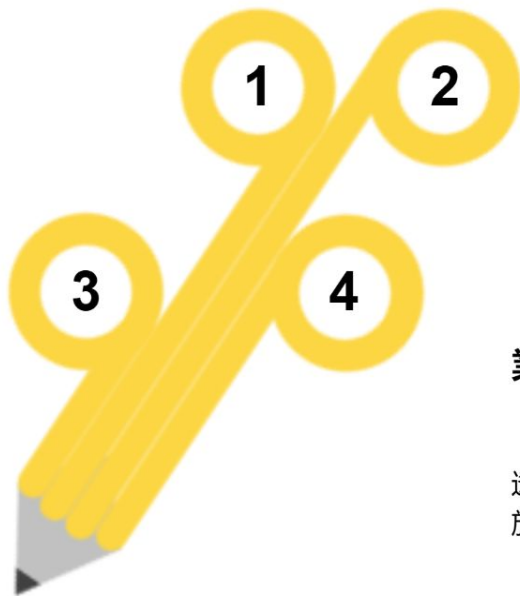
結合分析結果，幫助業務團隊找出高風險客戶群，以便進行針對性行銷

## 預警系統建議

建立高風險預警系統，及時介入減少客戶流失帶來的損失

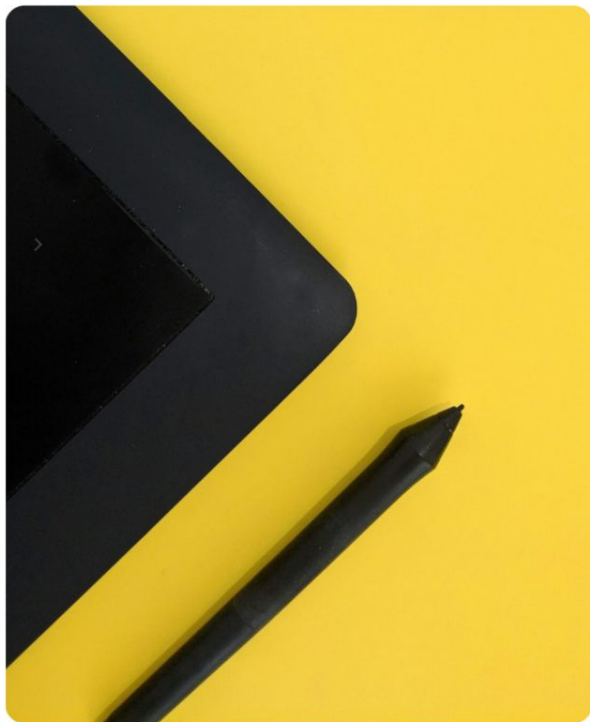
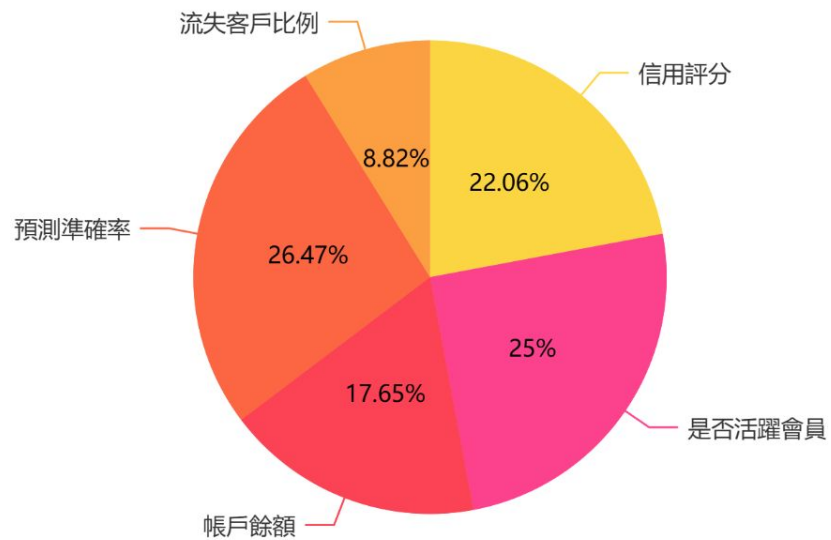
## 業務策略優化

透過保留獎勵機制提升客戶活躍度，聚焦於高CLV客群，增強獲利能力



# 銀行客戶流失預測分析

關鍵因素對於銀行客戶流失預測的重要性(百分比)





## 2 模型選擇與評估

# Logistic回歸與隨機森林模型在預測中的應用優勢

1

## 模型選擇的依據

選擇Logistic回歸作為基準模型，適合解釋性分析

2

## 隨機森林的優勢

隨機森林處理非線性與變數交互作用的能力更強大

3

## 模型性能比較

整體而言，隨機森林優於Logistic回歸，表現更佳

4

## 人員參與的重要性

數據分析師在模型選擇與評估中扮演關鍵角色



# 流失預測模型效果比較的關鍵指標及其解釋

## 評估指標概述

Accuracy、Precision、Recall和F1score是常見的評估指標

## 指標意義解析

各指標對於模型的有效性和適用性提供了不同視角

## 實際應用影響

高精度模型能有效預測客戶流失，提高商業回報

## 業務決策支持

模型的指標結果為業務決策提供了數據支持



# 如何利用混淆矩陣視覺化模型結果以便於分析

## 混淆矩陣的功能

幫助分析模型預測結果的正確性和錯誤類型

## 深入分析錯誤

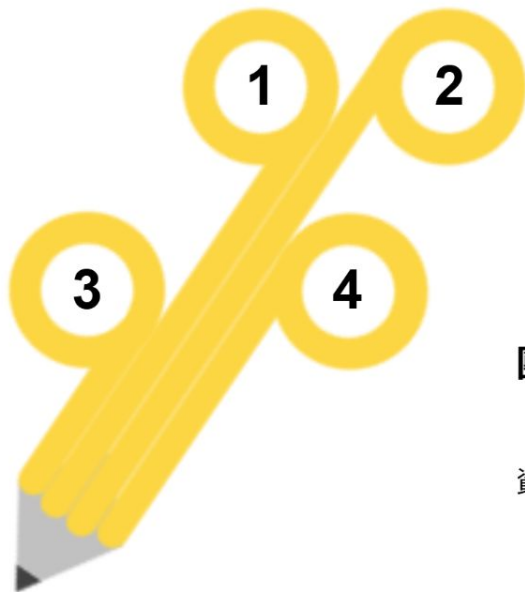
透過混淆矩陣可以針對錯誤預測進行深入分析

## 視覺化的必要性

清晰的視覺化能讓團隊更快速理解預測效果

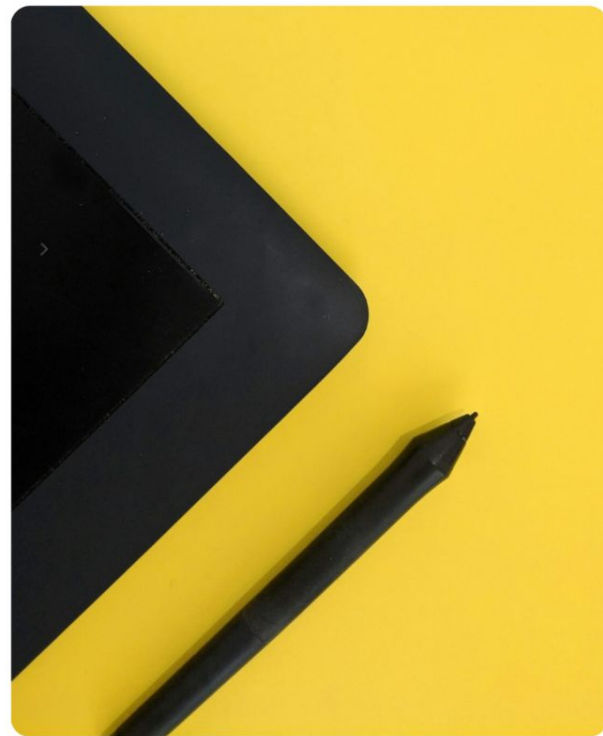
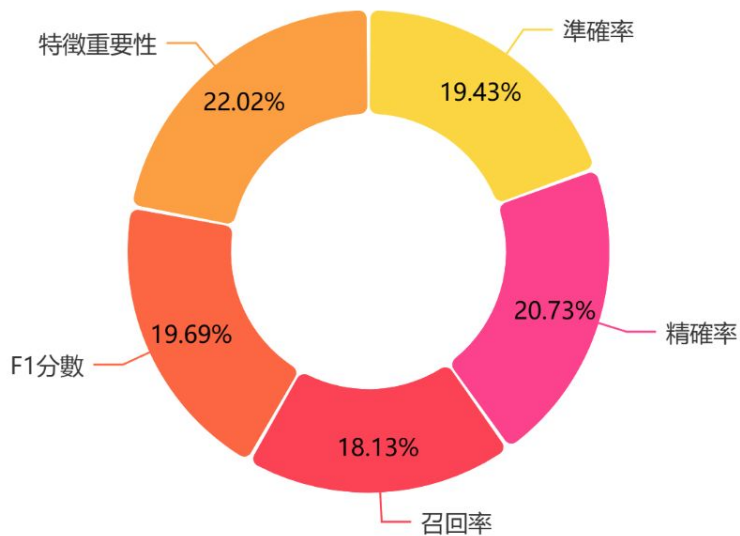
## 團隊協作的價值

資料視覺化助於促進團隊內的討論與決策



# 流失預測模型效果比較

Logistic回歸與隨機森林模型的效能比較(百分比)





# THANK YOU

Alex Huang