

嘖嘖募資專案之成效分析

G4

經濟四柯逸萱 B07303045

經濟四江于昕 B07303081

工管四陳薇守 B07701123

生工四邵宇彤 B07602019

生傳四盧家雯 B07610058

Agenda

01 . 研究主題

02 . 研究方法

03 . 研究結果

04 . 結論

PART 01

研究主題

研究嘖嘖募資平台的專案成效，有哪些重要的特徵或優勢能影響專案的成敗，並提供提案者在上架募資專案時，內容撰寫的注意事項之建議。

A teal vertical bar is on the left side of the slide. A teal L-shaped bracket is positioned to the left of the text 'PART 02' and '研究方法'.

PART 02

研究方法

- 網頁爬蟲
- EDA (探索式分析)

網頁爬蟲

- 資料來源：

本專案的資料來源為[嘖嘖募資平台](#)上的專案資料。資料集內容包含 2021 年 12 月 18 日前已結束募資的群眾集資（3095個）和預購式專案（749個），資料總數共 [3844 個專案](#)。

- 資料抓取與處理：

使用 [Beautiful Soup](#) 套件抓取網頁資料，將各項資料整理成一共[17個變數](#)，放入模型進行訓練，其中有些變數即是爬蟲下來的原始結果，有些會再經過運算處理。

- 特殊資料處理：含18禁內容之募資專案：

使用[Selenium瀏覽器自動化](#)套件Chrome Driver進行自動化點擊進入18禁專案。

資料欄位

依變項是**success**，募資成功（達到目標金額）為1，募資失敗為0，其他資料欄位皆為自變項

欄位名稱	欄位解釋	變數型態
success	已結束的專案，依募資結果分為 success/failure	依變數
category	專案類別，訓練模型時將各個類別都轉為dummy variable	自變數
raised money	募到的金額	
image_num	專案頁面裡的圖片數量	
video_num	專案頁面裡的影片數量	
website, yt, fb, ig	有附上社群連結的為1，沒有為0	
fb_like	臉書粉專的讚數	

set_num	專案頁面的商品組合數量	自變數
set_price	商品套組的價格中位數	
time_expected	第一個商品套組預計實現的日期扣掉專案開始的日期，時間單位為天	
sponsor_num	贊助人數（有購買專案商品的人數）	
target money	目標金額	
duration	專案進行時長，時間單位為天	
comment_num	留言數量（贊助者才可留言）	

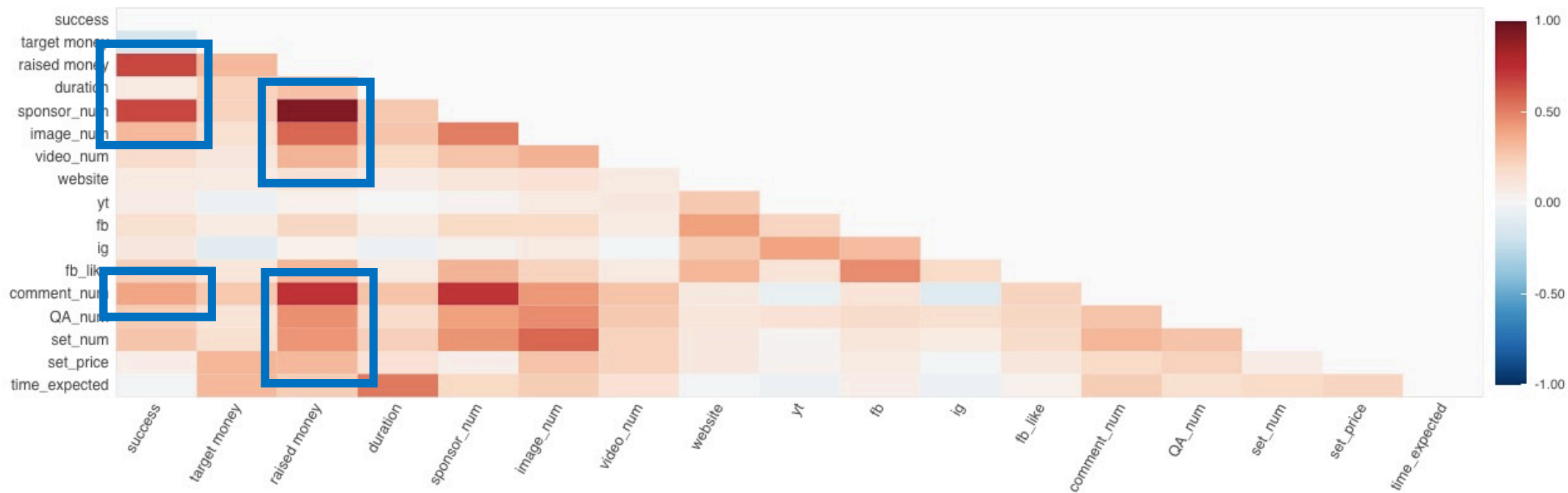
EDA：募資結果與專案分類

- 成功專案的總數大於失敗
- 「設計」與「科技」類佔大宗，約佔資料總數六成
- 失敗比例與該分類的總數成正比



已結束專案的募資結果（按分類）

EDA：變數相關性



- 成功/失敗：與募資金額、贊助人數高度相關，與圖片數和留言數中度相關
- 募資金額：與留言數高度相關；與圖片數、影片數、問答數、商品套組數中度相關，且各相關係數都更大
- 影片、圖片、商品多或互動熱烈都可提高募資成功機率

PART 03

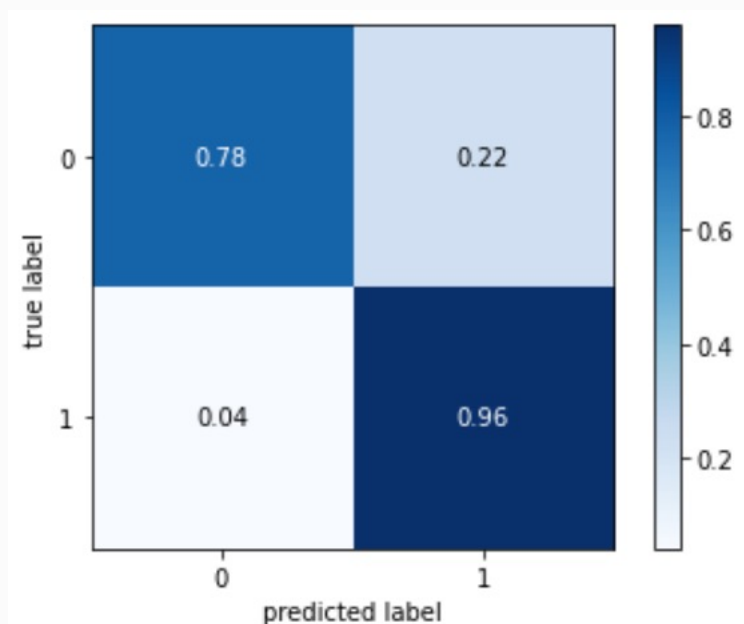
研究結果

- Unbalance v.s. Balance
- 事前預測 v.s. 在途預測
- Regreesor v.s. Classifier
- 測試資料集 v.s. 12/18-12/31真實資料集
- 成敗預測 v.s. 募資金額預測

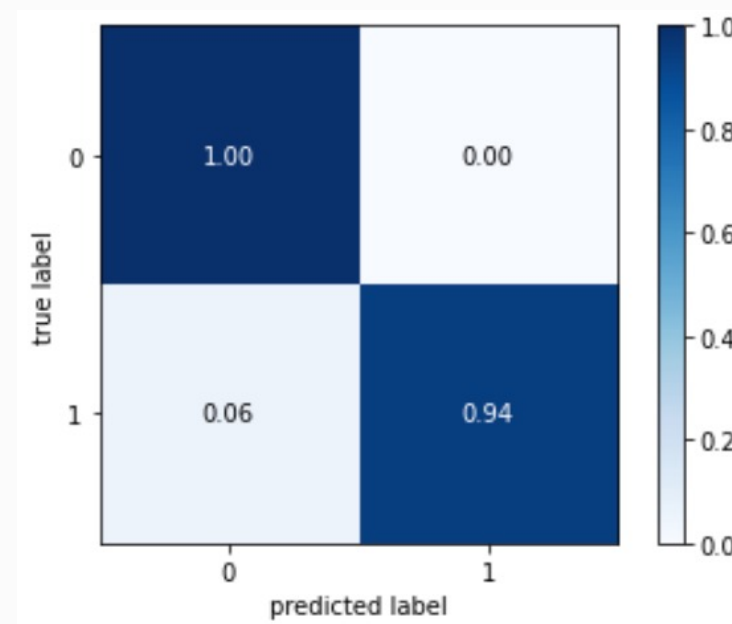
Unbalance v.s Balances

嘖嘖募資平台至今為止的**成功專案筆數為失敗專案的超過3倍**，不平衡的資料集使我們原先的預測結果更容易偏向1，由confusion matrix可看出balance之後1的準確率下降，但0的準確率提升許多，而**整體accuracy從0.92提升至0.97**。

資料筆數	Unbalance	Balance
1(成功)	3436	3436
0(失敗)	1128	3384
總筆數	4564	6820



Unbalance_Random Forest



Balance_Random Forest

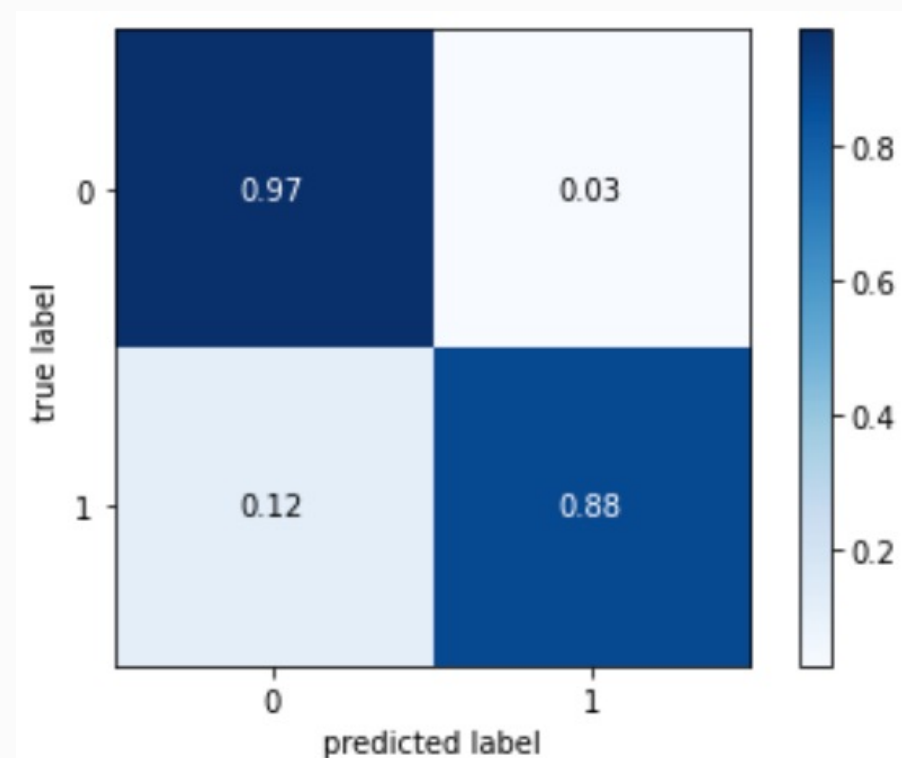
事前預測 v.s 在途預測

- 在途預測比事前預測表現更佳，Gradient Boosting的Accuracy 0.98為表現最佳之模型
- 事前預測中Random Forest的Accuracy 0.93為表現最好的模型

Accuracy	事前預測	在途預測
Ridge Regression	0.71	0.72
Lasso Regression	0.56	0.72
Random Forest	0.93	0.97
XG Boosting	0.75	0.94
Gradient Boosting	0.79	0.98

Note :

- 事前預測包括專案上架前有的參數，包括類別、圖片數、目標金額、募資時長等
- 在途預測除了事前預測的欄位外，再加上已贊助人數、留言數、QA問答數
- Ridge和Lasso以預測出來機率大於等於0.5為預測成功，小於0.5為預測失敗



事前預測_Random Forest

Regressor v.s Classifier

前十大特徵中，可以看出Regression多以「類別參數」的Coefficient為大，而Random Forest的前十大以「連續參數」為主，後者也是我們認為較符合直覺與常理的重要特徵。

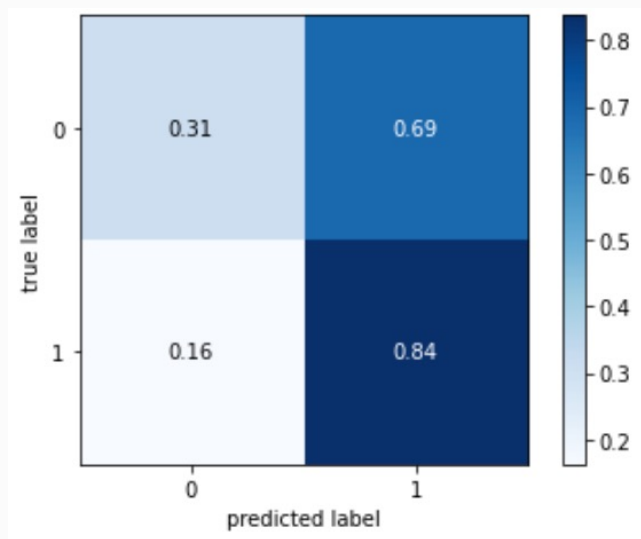
事前預測		在途預測	
Ridge Regression	Random Forest	Ridge Regression	Random Forest
挺好店	image	挺好店	sponsor_num
表演	fb讚	社會	comment_num
出版	target	表演	target
社會	duration	電影動畫	image
fb	time_expected	出版	fb讚
音樂	set_price	fb	set_price
電影動畫	set_num	音樂	time_expected
遊戲	video	教育	duration
ig	群眾募資	地方創生	set_num
時尚	fb	遊戲	QA_num

12/18-12/31結束的專案項目預測

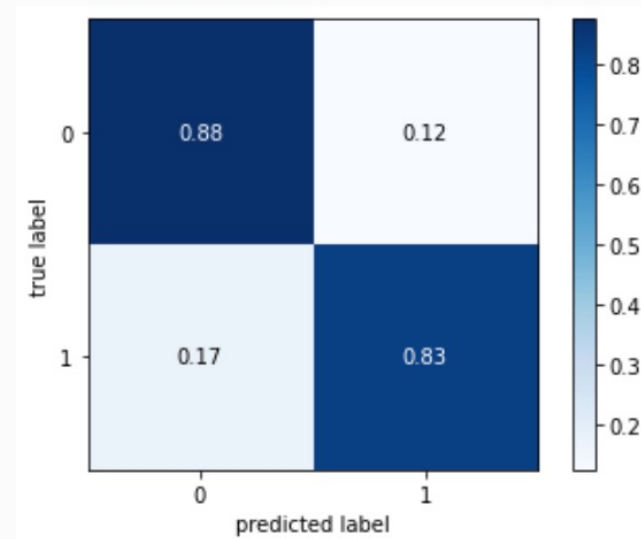
本組使用12/18日前的資料進行模型訓練(Random Forest)與預測，並在事後抓取12/18-12/31期間完成的專案來進行實測，共102筆資料，Accuracy不如原先的測試資料集。

推論：原本的訓練集進行過balance，因此測試的資料可能跟訓練的資料一模一樣，所以預測準確率被高估。

Accuracy	事前預測	在途預測
測試資料集	0.93	0.97
12/18-12/31 資料集	0.75	0.83



事前預測_12/18-12/31資料集



在途預測_12/18-12/31資料集

成敗預測 v.s 募資金額預測

本組除了使用分類器預測成敗，也有對於總募資金額進行預測，然而模型表現極差，RMSE皆為百萬起跳。募資金額之平均數為131萬，中位數為20萬。

此外我們也進行了PCA測試(Principal Component Analysis)，然而發現自變數對於money(依變數)的解釋程度不高，所以這也是我們認為模型表現不佳的主要原因。

- 推論與改進：
 1. 募資金額多寡除了跟專案的資金需求、目標金額有關之外，可能還受到贊助者願意出多少錢影響，但贊助者的願付金額高低在本模型中並無資料、也不易取得資料。
 2. 可嘗試的改進方式：將依變數改為「募得資金佔目標金額的比例」，把預測目標標準化。

PART 04: 結論

- 目前Random Forest為事前預測與在途預測綜合整體表現最好之模型，但Gradient Boosting於在途預測中表現極佳。待資料量更多時須持續試驗不同模型，確保最佳預測模型。
- 各類模型中並沒有對於「成功」或「失敗」的預測表現明顯偏頗，以Random Forest來說，事前預測測試資料集中失敗預測較精準、12/18-12/31資料集中成功預測較精準，而在途預測皆為失敗預測較精準。
- 綜合十大特徵與本組對募資平台的研究，「聲量」以及「熱度」為募資成功的關鍵要素，甚至「預熱」準備與策略都是決定募資成敗的關鍵，建議團隊必須有一定程度的行銷，包括製圖、拍影片以及社群媒體的經營都是必須注重的。