

1. 小組成員: 統計113 葉詠馨、黃筱云、郭旻霏

2. 競賽敘述及目標:

透過各種統計分析模型，針對不同特徵，預測客戶會不會離開這間銀行

3. 資料前處理:

先將無關的資料刪除，再將其他數據標準化，把名目資料轉成數值型資料

4. 特徵處理與分析: 將所有數據標準化

5. 預測訓練模型: K-means、KNN、DBscan、XGBoost

6. 預測結果分析:

剛開始我們都用較熟悉的統計模型進行預測，最後卻發現準確率並沒有到太高，之後用了另一個比較不熟悉的模型XGBoost，準確度才有提升。

History <small>The dataset you submitted</small>					
#	上傳時間	上傳檔案	Accuracy	Precision	FScore
6	2021-12-16 22:39:08	3.csv	0.8675	0.7	0.6131386861313869
5	2021-12-16 22:25:20	2.csv	0.795	0.41935483870967744	0.24074074074074076
4	2021-12-16 22:24:43	3 (1).csv	0.5325	0.1978021978021978	0.277992277992278
3	2021-12-16 22:20:27	upload2.csv	0.795	0.41935483870967744	0.24074074074074076
2	2021-12-16 22:19:01	2.csv	0.795	0.41935483870967744	0.24074074074074076
1	2021-12-16 22:07:30	1.csv	0.795	0.41935483870967744	0.24074074074074076

7. 感想與心得:

葉詠馨：老實說要不是有老師舉辦的這個競賽，我真的不會想主動去做這件事，多虧老師的安排，讓我趁著這個機會自己去學習。其實有很多時候我對於套件的內容還不是很了解，但透過這個機會多少有促使我去思考、理解。這個學期的課程蠻多、蠻繁重，雖然我真的沒有百分之百的完成每一個活動，而且一開始蠻受打擊的、都覺得自己寫不出來，但我真的覺得我有學到很多我沒有接觸過的東西，我會趁著暑假再利用moodle好好複習，更精進自己！

郭旻霏：對於我們來說這次的比賽其實是非常大的挑戰，我們並不是在資料科學的部分學得特別專精，甚至已經跟不太上老師的速度，因此剛開始我們都認為我們連第一次預測都做不出來，但最後在朋友的幫助下做出了我們的第一次預測時，真的非常有成就感，也越來越有信心，其實我們並不要求自己有多出色，反而是更希望自己能夠在這次的比賽中收穫點甚麼，很慶幸有助教和朋友的幫忙，讓我們在做預測時不至於太過打擊(因為一直出現error真的快瘋掉了!)我想這大概這次比賽就是老師想讓我們了解到在這堂課學習到的工具可以運用在甚麼地方。我並不是一個特別有天分的學生，在學習的過程中也感到非常艱難甚至一度想放棄退選，但我很慶幸自己有留下來的勇氣，而我也確實在這堂課學習到很多程式技巧，很高興自己有修完這

堂課，也相信這些對我的未來會有很大的幫助。

黃筱云：在這次的競賽中，自我學習占了很大的一個部分。我覺得任何一個人，都必須養成自學的習慣，自學，就是一種獨立學習，獨立思考的能力。為了使預測的資料變得更加精準，我們必須自行上網查詢合適的套件並研究其功能，但是我們知道的套件有限，而且對模型裡面的描述也不是都能完全看懂，所以有的時候就變成純粹把資料丟進模型內，而不了解模型背後的運作方式。透過競賽，我可以逼迫自己去學習新的方法，但由於近期的課業繁重，無法全心全意地研究各種機器學習的模型，我覺得有點可惜！經過這次的競賽，我體驗到了什麼是"寓教於樂"，因為比賽不僅能夠激發大家的好勝心也能夠讓我們學到更多專業知識，真是一舉兩得。

https://carrieyeh.github.io/DataScience/hw3_H24091265/Untitled-1.html

https://github.com/MiffyKuo/TA_class

https://miffykuo.github.io/TA_class/hw3/hw3.html.html

<https://github.com/SharonHuang04/Data-Science-HW>

<https://sharonhuang04.github.io/Data-Science-HW/HW3/hw3-html/index.html>

