應徵職位：資料分析工程師

求　職　者：林燕羚

- 學歷背景
- 個人介紹
- 專案介紹

學歷背景

2020 – 2022

國立陽明交通大學　統計學研究所

統計學助教 · 電機學院 / 數學系

論文 : 管制圖的迴歸預測法

- 中國統計學社 111年論文獎 – 佳作獎
- 110學年度 統計所獎學金

2015 – 2019

國立成功大學　數學系

- 107學年度 校際統計學競試 – 特優獎
- 107學年度 數學系 高等微積分學科獎學金

個人介紹

專業技能 ››› 了解統計檢定方法、統計領域和機器學習等相關知識。

專案經驗 ››› 有不同領域專案經驗，能跨領域溝通，並建立分析架構解決問題。

團隊合作 ››› 擔任協作者或溝通者的角色。

自主學習 ››› 線上課程、SQL、數據相關講座。

專案介紹

1. 管制圖的迴歸預測法 ·（品質管制）
2. 過動症患童之量化腦波分析 ·（機器學習）
3. 信用卡客戶流失預測 ·（機器學習）

專案一：管制圖的迴歸預測法（Thesis）

背景

- 生醫領域中數據取得不易，因此在一些儀器的管制/參考區間建構上有一定的困難。
- 某些製程中因干擾因素而產生的不穩定樣本，其建構出的管制界限會有誤差。

目標

- 研究一個新的管制界線建構方法來應對此狀況。
- 驗證該方法建構出的管制圖是否合理。

專案一：管制圖的迴歸預測法（Thesis）

方法

1. 使用 線性迴歸法 轉換 已知 的管制界限，並將其調整。
2. 將新方法運用不同參數狀況下來模擬管制界限。
3. 使用 平均運行長度（ARL）對於模擬結果進行評估。

結果

1. 模擬結果（ARL0）符合理論值370，表明了該方法的準確性和可行性。
2. 可推廣至不同形式的管制圖，也可應用在生醫的參考區間。

專案一：管制圖的迴歸預測法（Thesis）

舉例

- 找到一個函數 $G(x;\beta_0,\beta_1)$，使得 預測的管制界限 = 實際的管制界限。
- 以平均值管制圖為例，兩個監控的統計量 $\bar{X}$ 和 $\bar{Y}$，假設兩監控統計量具有線性關係，即

$$\bar{y}=\beta_0+\beta_1\bar{x}+\varepsilon=G(\bar{x};\beta_0,\beta_1)+\varepsilon$$

，β_0為截距參數，β_1為斜率參數，ε為誤差值。

	已知	實際		預測
	$\bar{X}\sim N\left(\mu_x,\frac{\sigma_x^2}{n}\right)$	$\bar{Y}\sim N\left(\mu_y,\frac{\sigma_y^2}{n}\right)$		$\bar{Y}$
管制上界	$UCL_x=\mu_x+3\frac{\sigma_x}{\sqrt{n}}$	$UCL_y=\mu_y+3\frac{\sigma_y}{\sqrt{n}}$	? =	$PUCL_y=G(UCL_x;\beta_0,\beta_1)=\beta_0+\beta_1UCL_x$
管制下界	$LCL_x=\mu_x-3\frac{\sigma_x}{\sqrt{n}}$	$LCL_y=\mu_y-3\frac{\sigma_y}{\sqrt{n}}$		$PLCL_y=G(LCL_x;\beta_0,\beta_1)=\beta_0+\beta_1LCL_x$

專案一：管制圖的迴歸預測法（Thesis）

透過 $(\bar{X},\bar{Y})$ 的聯合分佈，及其母體線性迴歸方程式，求得 $\beta_1=\sigma_{xy}/\sigma_x^2$， $\beta_0=\mu_y-\beta_1\mu_x$

但 $PUCL_y=\beta_0+\beta_1UCL_x=\mu_y+3\frac{\sigma_y}{\sqrt{n}}\rho_{xy}$ $\neq$ $\mu_y+3\frac{\sigma_y}{\sqrt{n}}=UCL_y$

$PLCL_y=\beta_0+\beta_1LCL_x=\mu_y-3\frac{\sigma_y}{\sqrt{n}}\rho_{xy}$ $\neq$ $\mu_y-3\frac{\sigma_y}{\sqrt{n}}=LCL_y$

經由調整，得出

新轉換預測法

$$PUCL_y=\frac{\beta_0+\beta_1UCL_x-\mu_y}{\rho_{xy}}+\mu_y\ ,\quad PLCL_y=\frac{\beta_0+\beta_1LCL_x-\mu_y}{\rho_{xy}}+\mu_y$$

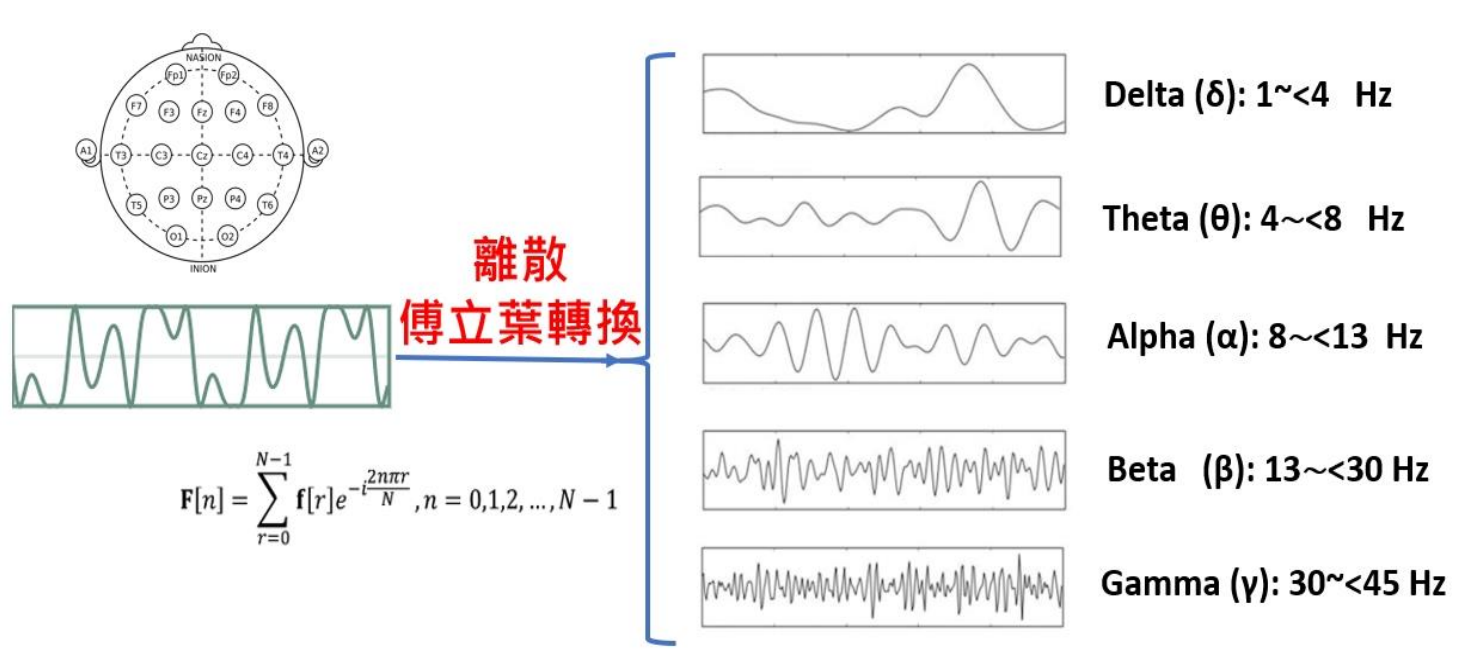
專案二：過動症患童之量化腦波分析

目標

根據 不同狀態 (休息狀態 - 睜眼/閉眼) 下的量化後腦波，能否客觀的判斷是否患有過動症 (ADHD) 。

數據

個人資料、多種量表紀錄及不同狀態下腦波數據等三面向資料，共129筆。



專案二：過動症患童之量化腦波分析

方法

1. 資料前處理
   - 數據清洗
   - 線性判別分析（LDA）假設：使用 Box-Cox Transformation，使其符合常態性。
2. 模型與訓練
   - 特徵編碼
   - 目標變量 比例（6：4）且 數據有限： 使用 分層交叉驗證 方法。
   - 使用 6 種機器學習模型來做分類預測。

專案二：過動症患童之量化腦波分析

評估

以 Recall（召回率）為主，F1 score 為輔。

	實際 YES	實際 NO
預測 YES	**TP** (True Positive)	**FP** (False Positive)
預測 NO	**FN** (False Negative)	**TN** (True Negative)

$$\frac{TP}{TP + FP}$$ **Precision**

$$\frac{TP}{TP + FN}$$ **Recall**

$$\textbf{F1 score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}$$

專案二：過動症患童之量化腦波分析

結果

	Box-cox transform + standardize			standardize		
Model	Recall	F1 score	Accuracy	Recall	F1 score	Accuracy
Logistic	0.85	0.87	0.85	0.83	0.85	0.83
KNN	0.74	0.80	0.78	0.71	0.80	0.77
LDA	0.75	0.79	0.77	0.59	0.62	0.61
SVM	**0.92**	0.90	0.88	0.83	0.82	0.79
Random Forest	0.85	0.85	0.83	0.83	0.84	0.82
XGBoost	0.83	0.84	0.81	0.83	0.84	0.81

專案二：過動症患童之量化腦波分析

結論

- 在此專案中，進行 Box-cox transform 後的各項指標普遍有變好。
- SVM 模型在 Recall 和 F1 score 的綜合表現最好
- 結果表明機器學習可實現對ADHD的客觀診斷，作為輔助醫療決策之參考依據。

改善

- 納入個人或量表的其他欄位資料或提高樣本量，來增加模型的預測能力。
- 嘗試使用更多的特徵工程方法，如 PCA 。

專案三：信用卡客戶流失預測

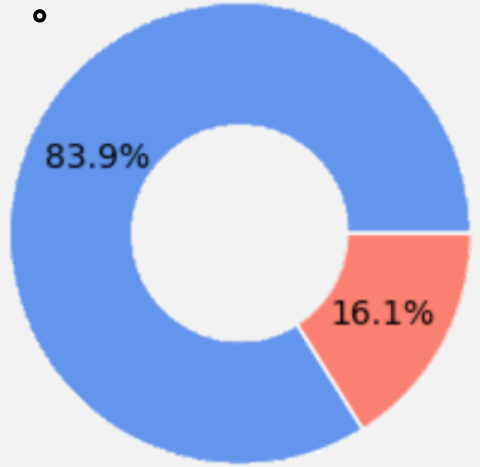
目的

透過相關數據，建立預測模型識識辨識潛在流失客戶，以便制定針對性的行銷策略和服務，增加客戶的忠誠度。

數據

信用卡帳款記錄、客戶特徵及往來關係等資料，共 10127 筆。

其中 留存：8500 筆（84%）、流失：1627 筆（16%）。



專案三：信用卡客戶流失預測

方法

1. 資料前處理
   - 離群值檢測
   - 相關性分析
2. 模型與訓練
   - 特徵編碼： Label Encoding、One-Hot Encoding
   - 模型： Random Forest
3. 模型優化
   - 不平衡數據處理： 使用 SMOTE 過採樣 方法。
   - 超參數優化： Grid Search

專案三：信用卡客戶流失預測

評估

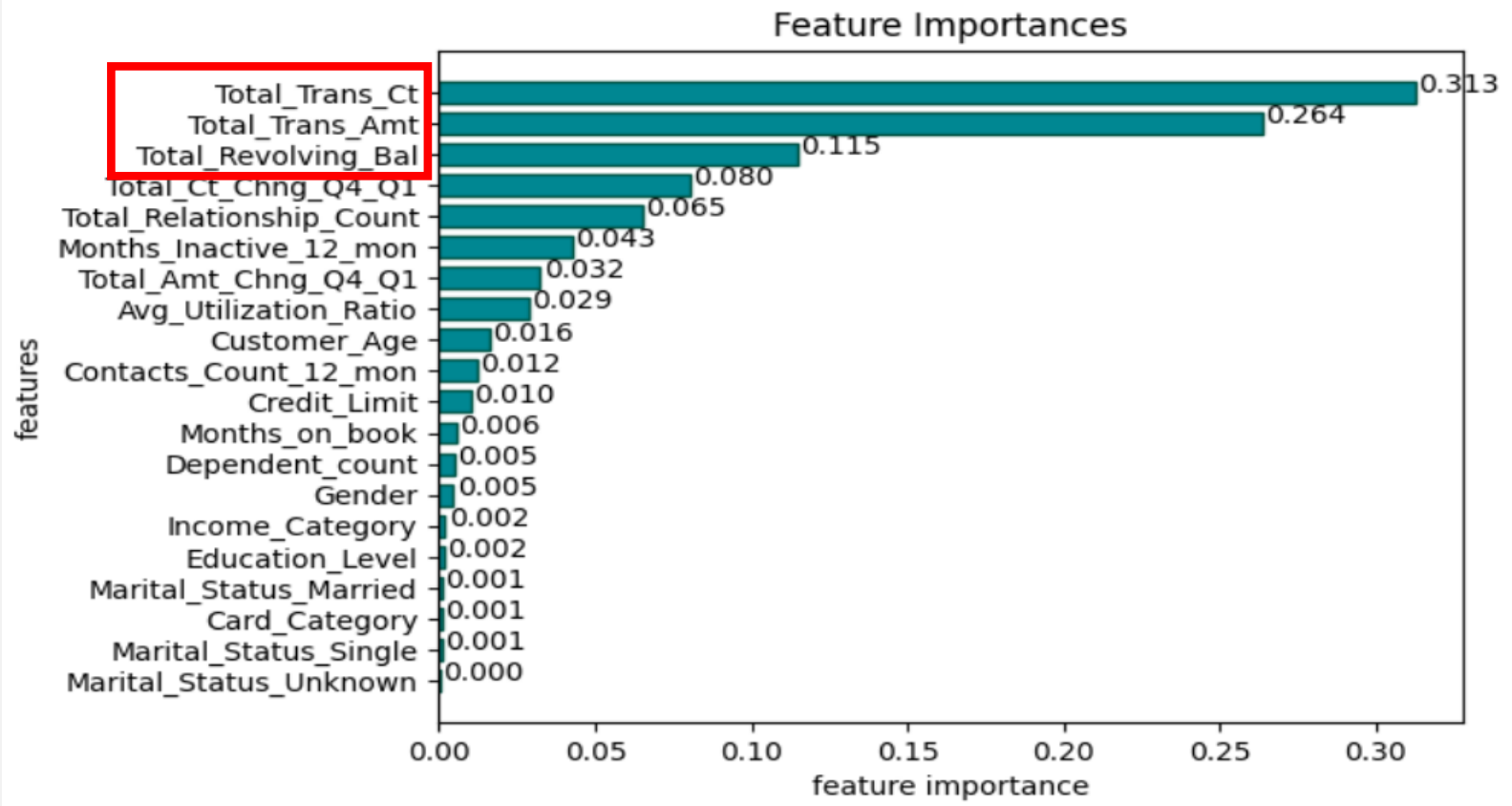
以 Recall（召回率）為主，F1 score 為輔。

結果

	Recall	Precision	F1 score	Accuracy
Basic	0.84	0.94	0.89	0.97
Basic + SMOTE	0.87	0.89	0.88	0.96
Basic + SMOTE + Grid Search	**0.91**	0.79	0.85	0.95

專案三：信用卡客戶流失預測

結果



Total_Trans_Ct -總交易數（過去12個月）、

Total_Trans_Amt -總交易金額（過去12個月）、

Total_Revolving_Bal -信用卡總周轉餘額。

專案三：信用卡客戶流失預測

結論

- 使用 SMOTE 能有效 提升 Recall，更好地識別潛在流失客戶。
- 重要特徵 與 信用卡交易 相關，可以此深入分析並提供針對性的行銷策略和服務。
- 整體而言，

  第 3 種方法 Recall 最高，但 Precision 最低；

  第 2 種方法 在 Recall 和 F1 score 均表現出色，同時保持較高的 Precision。

  在實務上，可根據業務需求和成本做權衡，選擇最適合解決方案。

THANKS