
金融科技期末專題

個人投資風險管家

— R09922129 詹鈞凱 R09631003鄭子揚 R09631017 吳少云 —

R09631033 陳玠宏 R09631050 詹閎棋

目錄

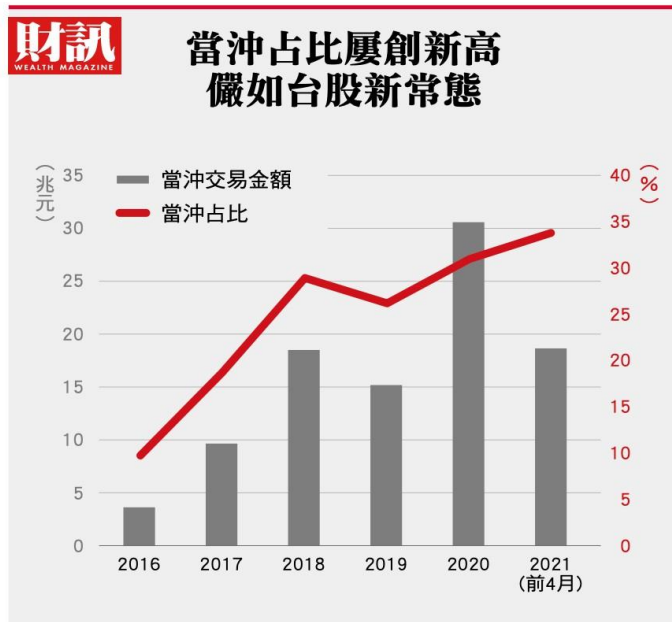
- 專案介紹
- 實驗方法與設計
- 實驗結果及討論
- 結論及展望

連結

- 短片
- 長片
- [原始碼](#)
- [小組開會記錄](#)

專案介紹

問題定義



資料來源：證交所

造成違約交割比例上升原因

- 投資人野心過大
- 戶頭可無現金下單
- 人為操作失誤

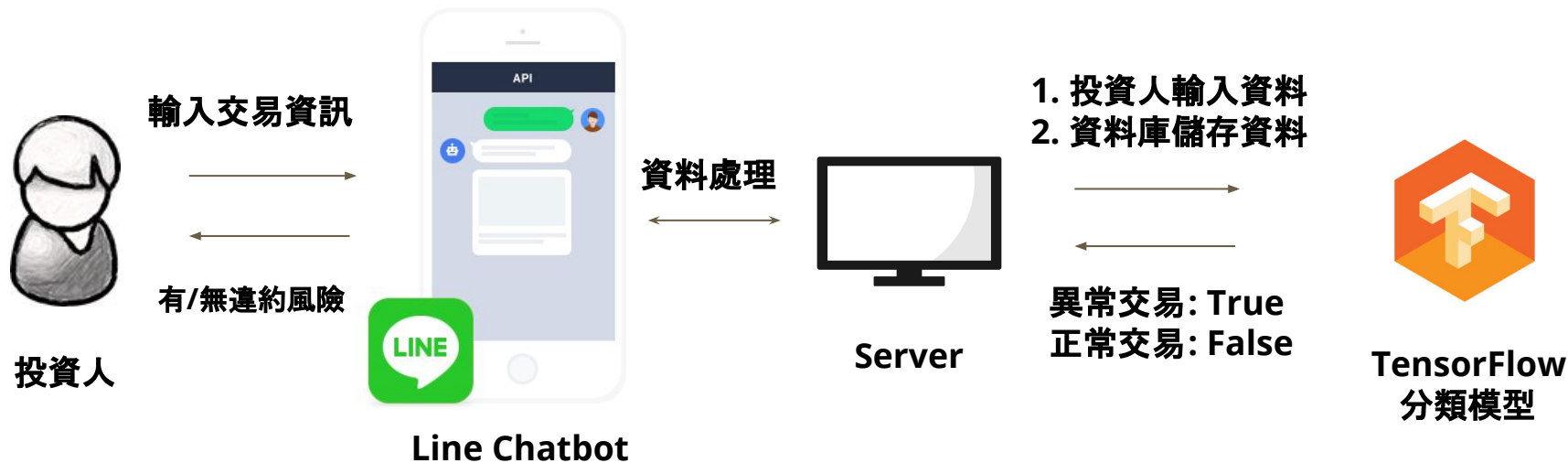
專案目標

投資一定有風險
基金投資有賺有賠
申購前應詳閱公開說明書

建立一套系統給予投資人下單的風險評估，作為客戶的投資風險管家

- 有效避免投資人的操作失誤
- 減少證券商處理違約交易的成本
- 提供意識投資風險的良好習慣

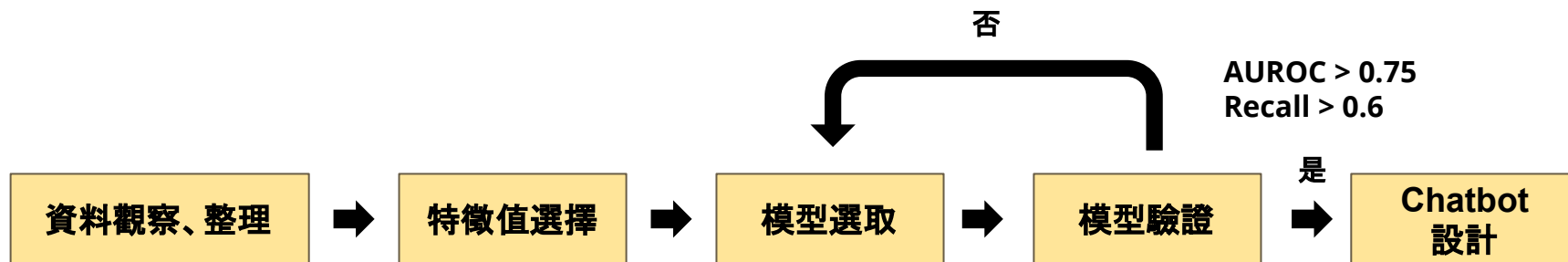
專案架構



專案Demo

實驗方法與設計

實驗路線圖



資料集

- **資料來源:** 玉山證券提供, 資料集時間: 2019 Q1 至 2021 Q1
 1. 2019 Q1至2021年Q1股票每日資訊, 如開盤、收盤價、當日最高、最低價格等
 2. 臺北股市股票之公司產業別
 3. 投資人有無違約紀錄, 包括 投資人年齡、違約註記、違約交易日等等
- **資料分析:**
 1. 有違約人數97, 無違約人數約100,000, **資料嚴重失恆**
 2. **違約記錄日期**不在交易資料日期範圍內
- **資料預處理:**
 1. 對**分布過大**的資料, 取log使其收斂到常態分佈例如: 平均交易股數
 2. **缺值**的部分依據違約的特性, 從原資料尋找合理數據進行補植

實驗設計 - 特徵值選擇

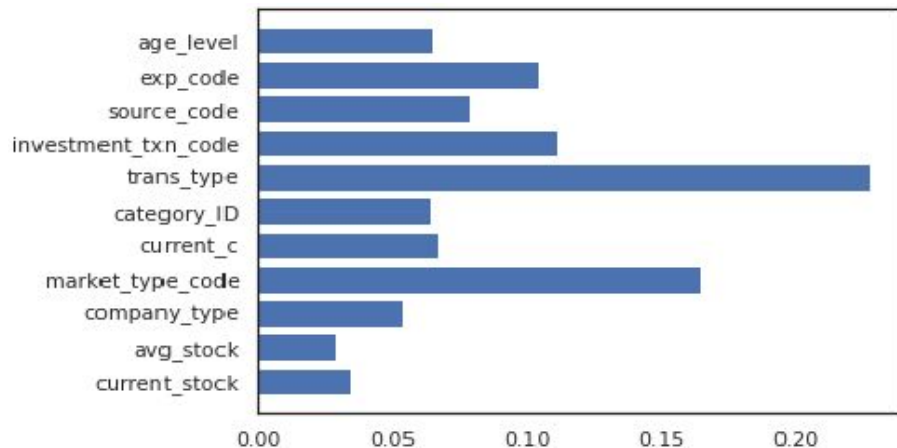
- 特徵依據來源可分為兩群

1. 當前交易的資訊 : 由使用者輸入
2. 過去市場消費經驗 : 統計各投資人過去的消費習慣並儲存於資料庫中

	特徵名稱	資料型態	備註
投資人輸入	投資者姓名	string	轉換成顧客代碼
	交易股票名稱	string	轉換成7種類股(current_c) 0: 電子, 1: 傳產, 2: 金融, 3: 食品生技, 4: 航運, 5: 其他, 6: No data
	交易股數(current_stock)	int	此交易的股票數量
	交易類型(trans_type)	int	0: 現股, 1: 融資, 2: 融券, 3: 現股當沖, 5: 借券, 9: 當沖
資料庫中資訊	年齡分布(age_level)	int	以年齡區間區分1~6之代碼 (玉證未提供詳細說明)
	交易經驗代碼(exp_code)	int	0: 未滿一年, 1: 1~2年, 2: 2~3年, 3: 3~5年, 4: 5~10年, 5: 10年以上
	開戶別(source_code)	int	1: 富國戶, 2: 一般開戶
	累積交易別	int	該投資人過去累積各交易類型的次數
	累積市場別	int	該投資人過去累積投資的公司類型
	平均持有股數	float	顧客以往交易的平均股票數量
	違約註記	boolean	0: 無違約, 1: 有違約

使用的特徵及資料型態

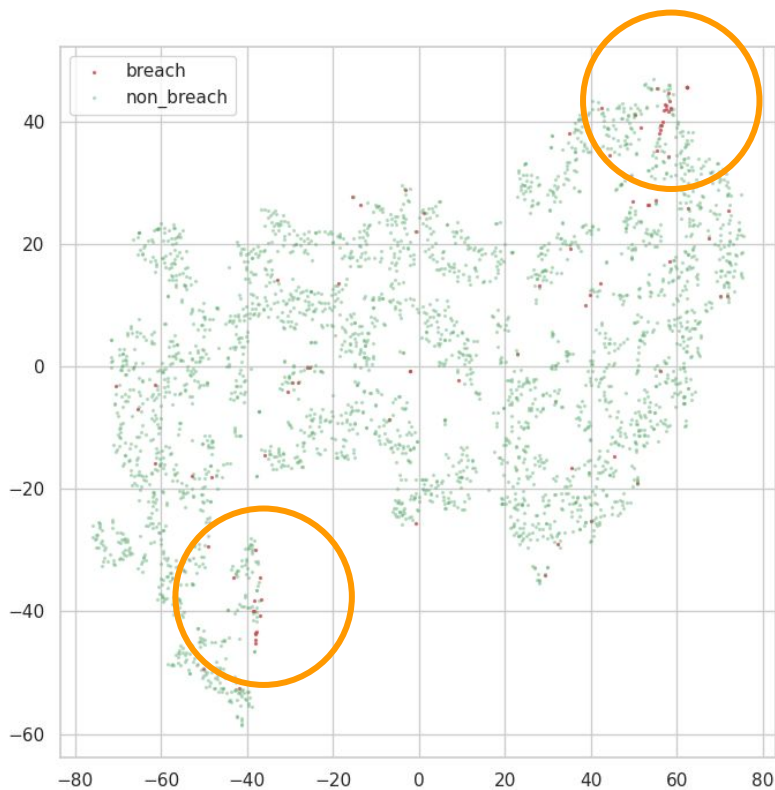
實驗設計 - 特徵值選擇



各特徵權重

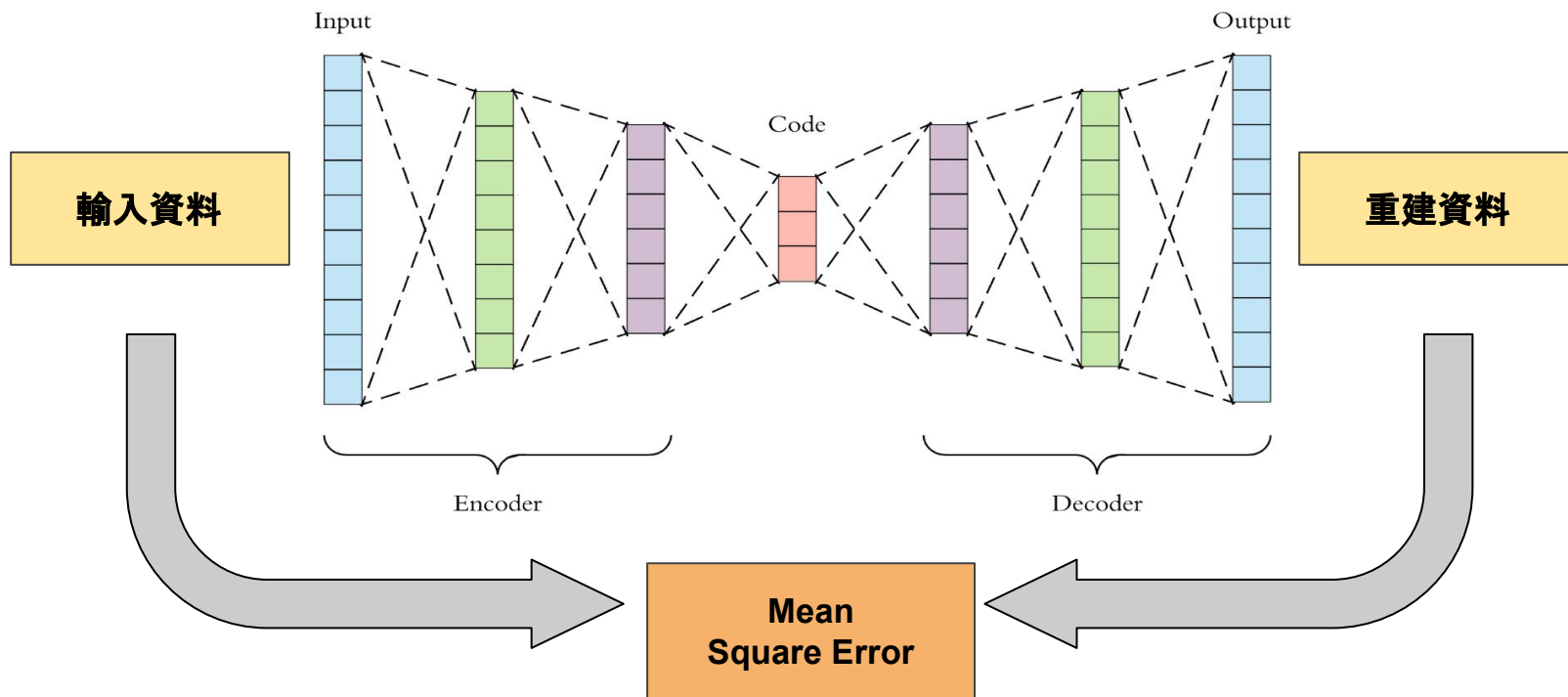
1. 將各支股票依據產業類別並劃分為六大類作為特徵
2. 使用XGBoost套件輸出個特徵重要性
 - 剔除低影響力特徵
 - 將影響力過高的幾項特徵刪除, 例如電子類股
以避免overfitting
3. 將投資人過去消費行為做累積(各類股、公司型態的投資次數)當作特徵

T-SNE資料二維分佈



違約分佈較為集中

實驗設計 - Autoencoder

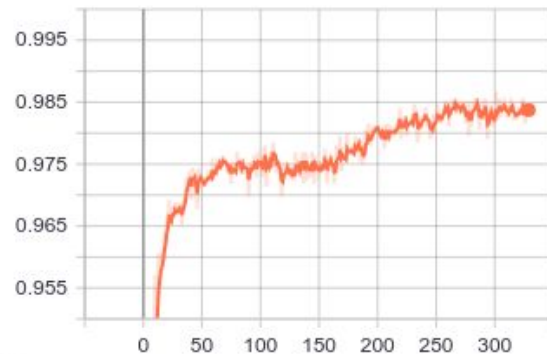


實驗設計 - 模型訓練

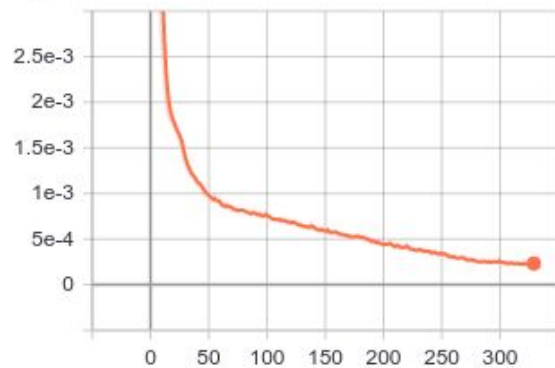
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 14)	210
dense_1 (Dense)	(None, 128)	1920
dense_2 (Dense)	(None, 64)	8256
dense_3 (Dense)	(None, 32)	2080
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 8)	136
dense_6 (Dense)	(None, 16)	144
dense_7 (Dense)	(None, 32)	544
dense_8 (Dense)	(None, 64)	2112
dense_9 (Dense)	(None, 128)	8320
dense_10 (Dense)	(None, 14)	1806
Total params: 26,056		
Trainable params: 26,056		
Non-trainable params: 0		

Batch Size = 128
Optimizer = Adam
Loss = MSE

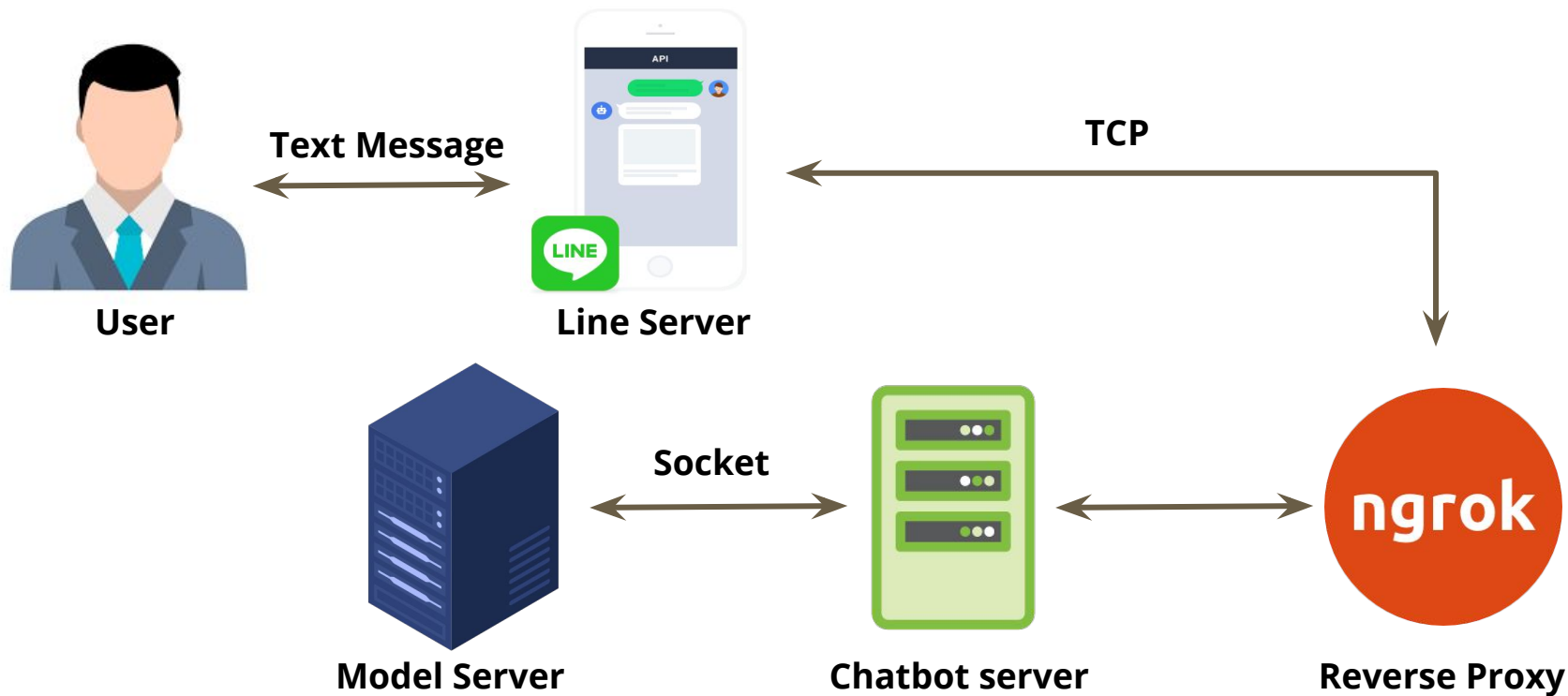
epoch_acc



epoch_loss

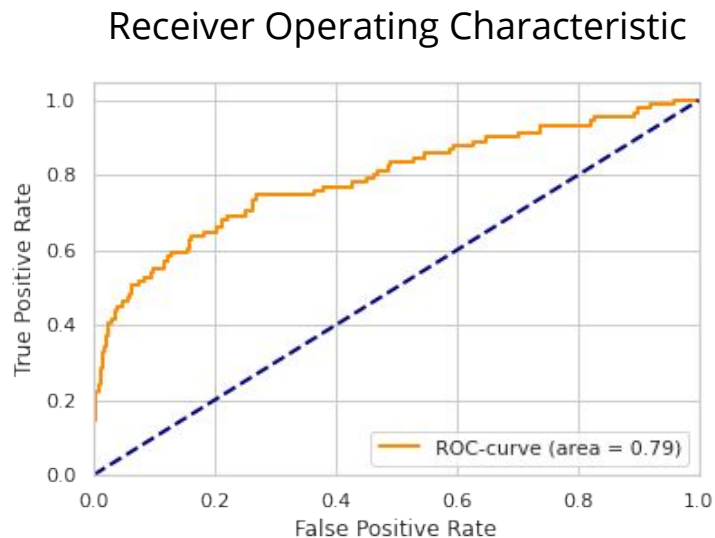


實驗設計 - Line Chatbot



實驗結果及討論

實驗設計 - 模型結果



True class	Predicted class	
	non_breach	breach
non_breach	433	87
breach	33	58

正常交易被正確判斷的比率 = $433/(433+87) = 83.27\%$

違約交易被判斷出來的比率 = $58/(33+58) = 63.74\%$

實驗結果 - 實例問題分析

Q：實際交易有違約但模型未判斷出來之情況

1. 交易紀錄正常，人為都難以判斷是否有異常

a. 使用General的交易紀錄所訓練的模型判斷難度高，須對該投資人的交易紀錄做客製化分類模型

avg_stock	current_stock
794.6666667	1000

2. 交易有明確異常，模型卻無法辨識，

a. 透過調整參數、加深模型，使其能夠學習地更完整

avg_stock	current_stock
27715.78947	47000

實驗結果 - Chatbot 實際畫面

實際操作步驟

1. 加入個人投資風險管家
2. 點選使用輸入說明，可顯示輸入資料格式
3. 下方「投資人」、「公司」和「類型」可更換輸入選項
4. **輸入範例「郭台銘,大立光,1000,0」**
5. 個人投資風險管家回傳此筆交易資訊

IF 異常: 回傳「!!! 此交易偵測為異常, 請再次確認!!!」

ELSE: 回傳「此交易未偵測到異常, 祝您交易愉快!」

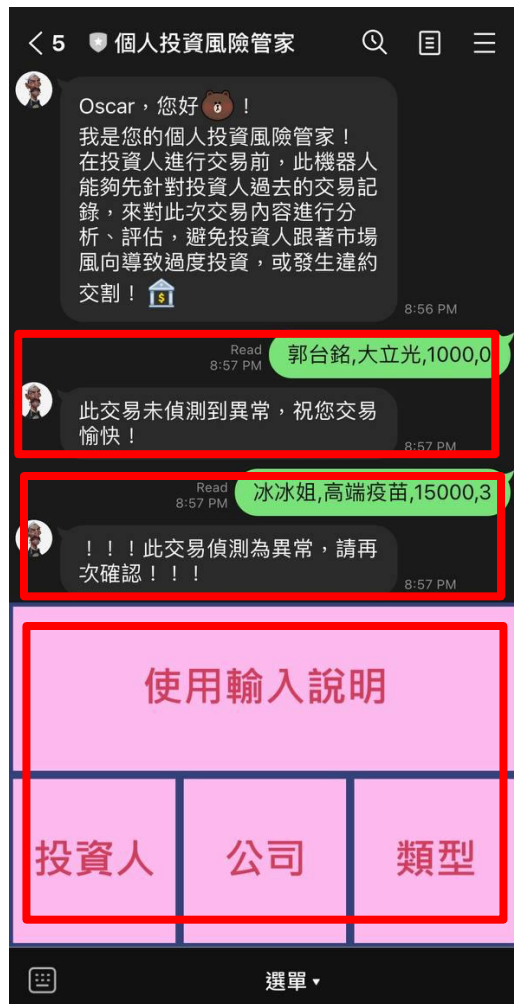
註: 本模型僅可用於玉山證券提供的歷史交易資料驗證, 如輸入範例中所列, 因此輸入不在清單的內容無法提供回饋

無違約結果

有違約結果

輸入說明及
輸入範例

Chatbot 實際操作頁面



結論及展望

結論與展望

- 針對**新開戶**使用者做額外的判斷
- 根據**單一投資人**過去的投資習慣對模型異常偵測結果做調整
- 根據每日交易資料，**滾動式調整模型**
- 根據每日交易資料做**股票熱度分析**，給予每筆交易多一項指標
- 使用RNN模型，將**交易順序及頻率**當作模型評估指標
- 將預測模型與玉山證券交易APP串接

附錄

參考文獻:

- <https://www.kaggle.com/robinteuwens/anomaly-detection-with-auto-encoders>
- <https://medium.com/@curiously/credit-card-fraud-detection-using-autoencoders-in-keras-tensorflow-for-hackers-part-vii-20e0c85301bd>
- <https://jason-chen-1992.weebly.com/home/-autoencoder>
- <https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>
- <https://developers.line.biz/zh-hant/services/bot-designer/>
- <https://ithelp.ithome.com.tw/articles/10192928?sc=hot>
- <https://xgboost.readthedocs.io/en/latest/>

原始碼: https://github.com/tabowsy/fintech_project

小組開會記錄: https://hackmd.io/hWgYN_TfRheM-0Mlf5AUQQ?view