

臺灣期貨交易所 114 年度研究報告提要表 填表人：邱鴻昌
填表日期：114 年 12 月 12 日

研究項目	機器學習應用於我國期貨市場異常交易行為偵測之研究		
研究單位及人員	臺灣期貨交易所 邱鴻昌、李浩宇、陳政朋	研究時間	114 年 1 月 1 日至 114 年 12 月 12 日

報 告 內 容 提 要

壹、研究內容重點：

本公司為順應全球金融市場的快速發展與數位化轉型浪潮，研議將機器學習技術應用於偵測「布假單」之異常交易行為，以提升市場監理效能。而鑒於期貨商品有數個不同到期月份契約、期貨商品價格差異甚鉅及市場資金輪動影響委託簿深淺等特性，本研究參考國外學者將非平穩的市場交易數據轉換為「擬平穩」的時間序列數據概念，嘗試以我國期貨市場所有期貨商品「成交單」作為特徵值轉換基準，透過以每筆成交前後之委託行為如委託價、委託量、委託時間、委託刪單價、委託刪單量、委託刪單時間及委託價、量變動影響委託簿最佳五檔揭示等面向來挖掘特徵，並將每筆「成交單」轉換成「特徵數值」，組成用於模型訓練及效能評估所需之資料集，再利用機器學習演算法進行異常交易行為偵測。

然期貨市場異常交易行為具多樣性，舉凡交易人透過彼此相互成交，達到其預期之交易目的，且該等相互成交其相互成交口數及比率極高，或伴有成交價格偏離或相互成交損益集中於特定人皆屬之，而該等行為具有尚難精確標註所有樣本屬於正常或異常之特性，故本研究另嘗試利用分群演算法區分正常交易與異常交易，並評估該技術之可行性。

貳、結論與建議事項：

本研究顯示，在本公司「正式環境」利用 Greenplum 資料庫結合機器學習函式庫 MADlib 之架構，確能建立完整的機器學習流程，且自資料準備、

特徵工程到模型訓練皆具備可行性，未來若進一步優化相關功能，應具有實務應用潛力。在方法論上，本研究提出若干具有實務應用價值之成果，包括跨商品的特徵正規化設計、分位數分層抽樣策略，以及整合市場監視經驗的特徵工程架構，可作為後續監理科技(SupTech)發展的基礎。

除監督式機器學習模型之偵測外，本研究另評估以非監督式機器學習偵測期貨市場其他疑似異常交易行為之可行性，先經降維、樣本權重調整，再以 K Means 演算法進行分群，於概念驗證階段發現能有效改善不平衡資料問題，提升分群效果。

本研究建構之布假單偵測模型在訓練階段表現良好，然用於預測新資料集之異常交易行為時精準度仍屬有限，主因係實務上真正的「真陽性 (True Positive)」案例比例極低，使模型訓練難度大幅提高，故建立乾淨、可追溯、具時間標記的案件資料庫，是機器學習能否有效發揮的關鍵前提，此與目前國外學者或國際監理機構看法發表之相關文獻結論一致。綜合國際發表的文獻及和國際各交易所交流經驗，運用機器學習偵測異常交易行為，仍未見成熟，似未有任一演算法成功適用於特定異常交易行為的偵測。建議未來精進方向應先從標籤流程優化、不平衡資料處理、特徵穩定性分析，以及混用多種建模演算法等面向著手，並納入專家智識反饋，以提升模型偵測之整體效能。

附註：一、報告內容提要應包括下列二部分：

- (一) 研究內容重點。
 - (二) 結論與建議事項。
- 二、本提要表須附電子檔