

AI股票預測回測報告

以下是一份專業且完整的小論文範例，專為應徵 AI 金融投資公司向管理階層簡報所設計。內容整合了你提供的 ryReport003_stock_prediction_gpu003.py.pdf、原始程式碼 (stock_prediction_gpu003.py, do_some_stats.py)，以及最新的回測結果 (all_backtest_summary.xlsx, model_performance_summary.xlsx)，強調技術架構、實驗流程、核心發現與產業應用價值。全篇以繁體中文撰寫，並於首次出現專有名詞時標註英文。

多股票高頻AI預測與量化回測：技術架構、數據實證與產業應用

摘要

本文提出一套針對美股 NASDAQ-100 (NASDAQ100) 多股票逐分鐘 (Minute Bar) 交易數據的**AI 量化投資研究框架**，結合**GPU加速運算 (GPU Acceleration)**、多種機器學習/深度學習模型，以及全自動化的資料處理、特徵工程與策略回測。系統不僅能自動遍歷多檔股票、多期數據與多模型組合，並可即時彙整多維度回測績效，利於學術、產業及金融工程應用。本報告將針對技術設計、流程、核心統計發現、策略績效及未來展望進行說明，並討論其作為 AI 金融投資產業落地的可行性與競爭優勢。

一、前言

隨著**高頻交易 (High-Frequency Trading, HFT)**、**量化投資 (Quantitative Investment)** 與**人工智慧 (Artificial Intelligence, AI)** 成為金融科技 (FinTech) 主流，傳統統計模型已難以捕捉市場微結構與短線機會。**機器學習 (Machine Learning, ML)** 與**深度學習 (Deep Learning)** 於股價預測及策略回測領域的應用逐漸普及，學術與產業界皆持續尋求更高效、更具泛化能力的 AI 金融系統。

二、系統架構設計理念

1. 多股票與大規模自動化

- 一次性批次處理**上百檔股票**，涵蓋多年度、多月份，支持產業級大數據規模。
- 自動遍歷各模型、特徵組合，利於橫向、縱向比較。

2. GPU 加速運算

- 深度學習模型（如 LSTM, CNN, Transformer）全面採用 GPU 運算，大幅縮短訓練與回測時間。

3. 特徵工程模組化

- 支援**價格平移特徵（Price Lag Features）、成交量平移、**多種技術指標（Technical Indicators, TA）與進階衍生特徵（Advanced Features）**動態組合。**

4. 策略與回測一體化

- 提供**全倉進出（All-in/All-out）、買進持有（Buy & Hold）、**定期定額（Dollar-Cost Averaging, DCA）**三大策略模組。**
- 各策略結果皆可自動批次回測並導出總結表。

三、資料流程與特徵工程

1. 資料來源與預處理

- **數據來源：**NASDAQ100成分股，每檔逐分鐘K線，涵蓋成交價（LastTradePrice）、成交量（Volume）、加權價（VolumeWeightPrice）、總成交筆數（TotalTrades）等。
- **預處理步驟：**轉換時戳（Datetime）、交易時段篩選、缺值補齊與排序。

2. 特徵工程（Feature Engineering）

- **平移特徵：**自動產生過去 N 期價格與成交量。
- **技術指標：**自動計算 RSI（Relative Strength Index）、MACD（Moving Average Convergence Divergence）、布林通道（Bollinger Bands）、EMA（Exponential Moving Average）等。
- **進階衍生：**如 VolumeValue、Min/Max Spread、時間特徵（小時/分鐘）等。
- **標籤設計：**預測下一期價格上漲(+1)或下跌(-1)，屬於二元分類（Binary Classification）問題。

四、模型設計與訓練流程

1. 傳統機器學習模型

- 線性回歸 (Linear Regression)：基線模型，評估線性關係。
- 隨機森林 (Random Forest, RF)：擅長捕捉非線性特徵與重要性排序。
- XGBoost (Extreme Gradient Boosting)：強化型決策樹，在金融領域具高準確度。

2. 深度學習模型

- 長短期記憶神經網路 (LSTM, Long Short-Term Memory)：專長時序建模。
- 一維卷積神經網路 (1D-CNN, 1D Convolutional Neural Network)：適合區段性特徵擷取。
- 簡易變壓器 (Simple Transformer)：具備全局序列依賴捕捉能力，適合高維序列資料。

3. 隨機基線 (Random Baseline)

- 設計隨機猜測 (Random Guess)，作為所有模型絕對表現之對照組。

4. 訓練細節

- 訓練/測試分割：測試集嚴格不混淆，確保時序合理性。
- 特徵標準化 (Standardization)：防止資訊外洩 (Data Leakage)。
- 評估指標：預測準確率 (Accuracy)、各策略回報率 (Return)。
- GPU自動偵測：若可用則自動切換至 GPU 運算，大幅提升效率。

五、投資策略設計與回測

1. 全進全出策略 (All-in/All-out)

- 根據預測結果全額進出，不使用槓桿，極端化驗證訊號價值。

2. 買進持有策略 (Buy & Hold)

- 以基準比較，被動型投資人常用方法。

3. 定期定額策略（Dollar-Cost Averaging, DCA）

- 每期等額買進，觀察平均成本與風險分散效果。

4. 回測結果匯總

- 每個【股票-特徵-模型-月份】自動導出準確率、策略回報、樣本數等。

六、核心實驗結果與統計發現

1. 統計總覽（以 2017 年 12 月為例）

- 股票數量：110 檔
- 模型數量：6 種（LinearRegression、RandomForest、XGBoost、LSTM、CNN、Transformer）
- 特徵組合：「價格+成交量+技術指標+進階特徵」

2. 各模型表現（平均準確率與策略回報）

模型	平均準確率	AllinOut均報酬	勝過Buy&Hold	勝過DCA	最高AllinOut報酬
LinearRegression	0.572	4.0%	104	105	21.8%
RandomForest	0.548	2.5%	96	97	23.9%
LSTM	0.552	1.4%	91	93	19.8%
XGBoost	0.528	1.8%	93	99	22.7%
CNN	0.527	0.04%	59	77	8.8%
Transformer	0.521	-0.25%	62	65	6.8%

- Buy & Hold 基準：各模型平均約 -0.6%。
- DCA基準：全體均值約 -0.46%。

3. 勝率與特例

- **LinearRegression** 勝率最高，95% 檔案勝過 Buy&Hold，且穩定性最佳。
- **最佳個股** (SIRI, RandomForest)：單月 All-in/All-out 策略報酬高達 23.9%，Buy&Hold 同期僅 1.5%。

4. 關鍵現象解析

- AI 模型整體能顯著提升投資回報，並有效擊敗傳統被動策略於下跌市況的表現。
- 線性組合模型於多特徵情境展現高穩定性，非線性模型（如RF、LSTM）適合捕捉特殊個股結構。
- 極短期回測的年化回報僅具理論意義，實務應以總區間報酬率為主。

七、程式自動化、數據彙整與視覺化

- **自動聚合與分組統計**：利用 pandas 及群組聚合方法，按模型計算平均、標準差、最大/最小報酬與準確率。
- **視覺化成果**：以 matplotlib 產製回報率與準確率長條圖、箱型圖，輔助管理層判讀策略優劣。
- **中英文對照摘要**：程式自動產出多語版本摘要表，利於國際化簡報或跨國溝通。

八、產業應用潛力與優勢

1. 研究與教學

- 作為金融工程、AI/ML課程教材，涵蓋從資料、特徵、模型到策略回測的全流程。

2. 量化基金與自營部門

- **量化對沖基金** (Quant Hedge Fund)、**券商自營部門**皆可利用本系統做新策略驗證及自動化開發。

3. 金融監理 (RegTech)

- 可協助監理單位驗證 AI 在資本市場的可控性、解釋性與風險評估。

4. 可擴展性與效能

- 模型與特徵組合彈性擴充，能快速調整應對多市場、多時段資料。
 - **GPU 並行加速**突破傳統 CPU 瓶頸，適合大數據與高頻場景。
-

九、技術挑戰與未來展望

- **非平穩性 (Non-Stationarity)**：須定期重訓以因應市場結構變化。
 - **過度擬合 (Overfitting)**：建議正則化與交叉驗證技術防範。
 - **實盤落地挑戰**：回測未考慮滑價 (Slippage)、手續費與流動性風險，需後續實測優化。
 - **多因子擴展**：建議導入多因子 (Multi-factor) 模型、Alpha 組合優化與風險因子分析，提升實務應用層級。
-

十、結論

本架構結合金融工程、資訊工程與 AI/ML，實現了高效的**多股票高頻預測與量化回測全流程自動化**。從資料處理、特徵工程、模型訓練到投資策略回測，皆具備高度擴充性與產業應用潛力。**不僅能顯著提升投資組合報酬與多元性，亦有助於未來台灣乃至全球量化金融創新與發展。**

如需完整數據表格、程式碼細節、PPT簡報或中英文對照，皆可進一步提供。

附錄：部分核心統計與程式碼段（可視需求增補於簡報）

若需進一步強化產業落地層面或補充特定模型（如 Transformer）/ 個股深度分析，可指定補充。
如需將統計圖表自動嵌入或編寫英文簡報版本，也可告知。