

2025/10/20、11/05

20:30 - 22:00

ML Group 3rd Meeting Minutes

Topic	checkpoint discussion
Place	Google meet
bullet points	<p>分享各自遇到的問題，列舉如下：</p> <ol style="list-style-type: none">1. 準確率太低：都在 50%左右，跟盲猜差不多的機率。2. 可能原因：使用的模型太簡單、特徵工程做得不夠好、股價漲跌滯後性與隨機性的影響。3. 滯後性：模型預測的股市走向相比於實際股市走向具有一定的滯後性，可能原因是透過該股票過去的價格與資料去做預測，而忽略了其他影響股價的因素，例如美股的波動或其他經濟因素對於該股票的影響。4. 包含絕對價格資訊的 feature(開盤價、五日均價等)可能對模型預測會有不利影響，像是超過某個價格之後，模型就認為不可能再漲。將這些 feature 轉換成相對指標，像是用收盤價是否高於五日均價，取代直接用五日均價作為特徵去訓練可能是一個解方。 <p>訓練台積電歷史股價資料，預測隔日漲跌：</p> <p>各模型初次預測結果簡述：</p> <ol style="list-style-type: none">1. KNN：最佳 K：26，準確率: 0.5323。2. Naive Bayes：在未對特徵與超參數做處理的情況下準確率約 0.5208，進行 fine tuning 後準確率約 0.5294。3. LSTM：使用 LSTM 預測實際股價，MAPE 落在 3-4%，但是對於漲/跌的預測準確率僅有 53%左右4. XGBoost：預測實際股價，RMSE 約 80.1，MAPE 則約 5%5. Random Forest：預測隔日是漲或是跌(平) 的二元分類，整體預測結果準確率為 52%，其中模型掌握了 76%的跌卻僅預測正確 27%的漲，表示模型比較會猜跌比較準，或是模型大部分時間都猜跌。

In attendance	郭宣汝(Present), 高予謙(Present), 謝尚錡(Present), 曾煜展(Present), 郭子維(Present), 盛樺(Present)
Task Assigned	<p>共同處理：</p> <ol style="list-style-type: none"> 1. 增加特徵：尋找除了技術指標外其他可能影響特徵，如外資買賣超等。 2. 降維處理：用 PCA 去掉冗餘特徵或用 Autoencoder 提取低維非線性特徵等。 3. Feature selection: <ul style="list-style-type: none"> 甲、Filter Methods 乙、Wrapper Methods 丙、Embedded Methods <p>各自工作：</p> <ol style="list-style-type: none"> 1. 模型測試與調參：預測 1 天／5 天後漲跌準確率，努力將 Precision 提升到 60%以上。 2. 實驗結果記錄：紀錄改進過程、準確率如何提升、自己對於準確率提升／下降與該模型的特性關聯的思考等。 3. 郭宣汝 :KNN、高予謙 :Logistic regression、盛樺 :Naïve Bayes、謝尚錡 :Random Forest、郭子維 :XGBoost、曾煜展 :LSTM。如若準確率受模型複雜度限制，可以選擇其他模型嘗試。如：CNN 等。
Next meeting	<p>Date: 2025/11/12</p> <p>Time: 20:30</p> <p>Objective: 分享模型如何改進和遇到的問題、怎麼解決問題。</p> <p>Location: Google Meet</p>