

1.

	Window size	Step	Train loss	Val loss	Best Val loss	Test loss
Original	10	15	228.6929	322.1089	322.1089	561.1099
First	10	5	7.0640	7.9724	7.9135	20.1268
Second	5	15	148.9350	146.5053	146.5053	435.5264
Third	5	5	5.7399	8.1836	8.1651	17.3576

根據以上結果，Window size 及 Step 越少時模型結果越佳，特別是 Step 對模型結果的影響更顯著，在 Window size 為 10 時，Step 由 15 調整為 5 後，Train loss 及 Val loss 由 200~300 下降至 7~8 左右，Test loss 也由 561 下降至 20，但所有結果 Test loss 相較於 Train loss 及 Val loss 都顯著較高，說明模型有 Overfitting 的情況發生。

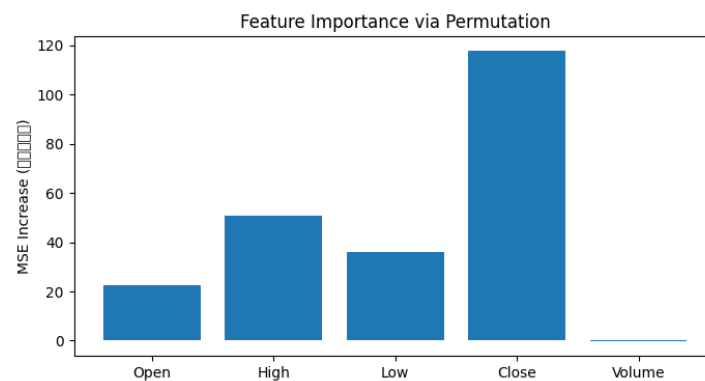
2.

(i)

	Window size	Step	Train loss	Val loss	Best Val loss	Test loss
Exclude 'Volume'	5	5	5.7399	8.1836	8.1651	17.3576
Include 'Volume'	5	5	869.5004	928.7477	928.6604	1136.3813

原模型在 Window size = 5、Step = 5 時，加入 'Volume' 這個 feature 後，Train loss、Val loss、Test loss 都顯著上升，主因可能為數值範圍過大，未經過標準化或正規化處理，亦有可能為 Volume 對股價的預測並無關聯。

(ii)



features	Window size	Step	Train loss	Val loss	Best Val loss	Test loss
'Open', 'High', 'Low', 'Close', 'Volume'	5	5	869.5004	928.7477	928.6604	1136.3813
'Open', 'High', 'Low', 'Close'	5	5	5.7399	8.1836	8.1651	17.3576
'High', 'Low', 'Close'	5	5	4.6299	6.7287	6.6981	14.6047
'High', 'Close'	5	5	4.4364	6.0670	6.0301	14.7538
'Close'	5	5	4.5313	6.3791	6.2099	14.1984

根據不同 feature 對 MSE 的影響由低到高逐一刪除，刪除 'Volume' 會使結果顯著由 1000 左右降低至 20 以下，在 feature 剩下 'High', 'Low', 'Close' 後結果都相近，feature 僅包含 'Close' 時模型結果最佳，太多 feature 可能導致有 noise 使模型結果較差，特別是 'Volume' 加入後反而導致模型結果變差。

3.

Before / After Normalization	Window size	Step	Train loss	Val loss	Best Val loss	Test loss
Before	5	5	869.5004	928.7477	928.6604	1136.3813
After	5	5	5.6384	9.0922	9.0748	15.5265

根據 2.(i) 結果加入 'Volume' 此 feature 會導致 MSE 的結果差異很大，可能原因為數值過大、無經過 Normalization，將所有 feature 都進行 Normalization，會使 Train loss、Val loss、Test loss 都顯著降低到 5 ~ 15 左右，結果說明 Normalization 會使模型預測結果更準確，避免部分 feature 主導模型學習、提升梯度下降的效率。

4.

window size 要小於 step size 因為在分割數據時要避免數據過度重疊、重複處理相同的數據，可以保持不同數據樣本的獨立性，比較小的 window size 還可以更準確的提取局部細節。

5.

Time Warping 為時間序列資料可以進行 data augmentation 的方法，透過伸展或壓縮時間軸上部分序列，讓時間序列產生平滑變化，並保留資料原始結構，可以模擬市場在不同時間內可能出現的不同變化，有助於模型學習在市場節奏改變下仍能穩定預測股價的能力。

6.

(i) Convolution-based models:

Convolution-based models 通常以固定大小的 window size 對輸入資料進行特徵擷取，在推論時 window size 大小會影響輸入張量的維度，輸入長度必須等於訓練時的 window size。較小的 window size 可提高局部特徵的分辨率，而較大的 window size 可涵蓋更廣的範圍。

(ii) Recurrent-based models:

Recurrent-based models 對於時間序列具有天然的時間依賴性，推論時 window size 大小可以動態調整。常見做法為使用 sliding window 或批量處理來確保長序列能被有效學習，避免梯度消失問題。LSTM 或 GRU 也可以透過隱藏狀態來維持長期記憶。

(iii) Transformer-based models:

Transformer-based models 架構通常使用固定長度的 window size 來分割輸入數據，並透過 Self-Attention 獲取全局關聯。推論時 window size 影響計算成本與記憶需求，較小的 window size 有助於減少計算負擔，而較大的 window size 可提高上下文理解能力。