問題 1:
如果要達成「能夠預測指定的 symbol 在 90 天後，是否有成長 10%」的目標，會選用的模型及訓練方式，文字提供選用的模型及原因

由於這是一個二元分類問題。所以我會在 Logistic Regression、Random Forest 和 neural network 之間進行選擇，考慮到資料是 High-dimensional data，因此我認為使用 neural network 是較佳的選擇，實際上，在股價分析中，Neural Network 通常能夠提供較高的準確性和預測能力。
在決定要使用哪種 Neural Network 時，我會在 LSTM 和 ResNet 之間作抉擇，最終，我選擇 LSTM，因為它能夠記住較長時間跨度內的歷史數據，這在處理時間序列數據方面具有獨特的優勢。
訓練方式方面，首先，由於數據已經過特徵工程處理，我會將數據進行標準化或正規化處理。接著，我會將數據集劃分為 Training Set、Validation Set 和 Testing Set，比例為 8:1:1。這樣的比例適用於數據量不夠龐大的情況。如果資料量非常大，則可以考慮使用 98:1:1 的比例。在 LSTM 模型的構建過程中，由於是二元分類問題，我會使用 Binary Cross-Entropy Loss 作為 loss function。在中間層，我會使用 ReLU 作為 activation function，這有助於模型學習非線性特徵。在輸出層，我會使用 Sigmoid 作為 activation function，將輸出轉為 0 和 1 之間的概率值。

問題 2:
在訓練中如何從現有的資料集提取出關鍵影響欄位

首先，必須納入 Date，因為它是時間序列資料的基礎，能夠幫助我們識別數據的時間順序和趨勢。接著，需要考慮包括 Open、High、Low 和 Close 等基本價格指標，因為這些數據直接反映了股票在每個交易日的價格波動情況。
此外，一些常見的技術指標如 williams、rsi 和 MA 也應該納入考慮範圍內，這些指標能夠揭示出股價的相對強弱、移動平均和超買超賣情況，對預測股價變動有很大幫助。Volume 也是重要的特徵，因為它代表著這支股票當前的市場交易活躍度和投資者情緒。
同時，standardDeviation 也是一個關鍵欄位，因為它反映了股價的波動程度，可以用來評估市場的風險。將這些欄位選入後，需要對數據進行標準化或正規化處理，這樣可以確保不同特徵的數值範圍相近，避免某些特徵對模型的影響過大。接著，可以使用相關性分析來篩選對預測影響較大的特徵。例如，計算每個特徵與目標變量之間的相關係數，選擇那些相關性較強的特徵，並剔除相關性較弱或冗餘的特徵。這樣，可以提高模型的訓練效率和預測準確性，從而更有效地達到預測目標。

問題 3:
如何利用目前已有的資料集欄位，推論出更有效的新資料欄位

可以使用 Heatmap 來查看特徵之間的相關性，這有助於識別哪些特徵之間存在高度相關性。接下來，可以組合高相關性的特徵來創建新的特徵。這樣做不僅可以降低模型的維度，還能提高訓練速度。例如，將價格和成交量相乘來獲得一個新的特徵，這個特徵能夠綜合反映價格和成交量的影響。