**Báo cáo LSTM cho Dự Báo Giá Cổ Phiếu Tesla**

**1. Lý do sử dụng mô hình LSTM:**

* **Xử lý dữ liệu chuỗi thời gian**: LSTM (Long Short-Term Memory) là một loại mạng nơ-ron hồi tiếp (RNN) đặc biệt hiệu quả cho dữ liệu chuỗi thời gian nhờ vào khả năng ghi nhớ thông tin trong thời gian dài và tránh vấn đề biến mất gradient.
* **Phù hợp với dữ liệu tài chính**: Mô hình này có thể xử lý sự phụ thuộc dài hạn trong dữ liệu giá cổ phiếu và các biến ngoại sinh như chỉ số S&P 500 và NASDAQ (IXIC), giúp cải thiện độ chính xác dự báo.

**2. Các kiểm thử cần thiết khi chạy mô hình:**

* **Phân chia dữ liệu**: Chia dữ liệu thành tập huấn luyện và kiểm tra (thường là 80-20) để đảm bảo mô hình được huấn luyện tốt và kiểm thử tính hiệu quả trên dữ liệu chưa từng thấy.
* **Kiểm thử hiệu suất**: Sử dụng các chỉ số như RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), và MAPE (Mean Absolute Percentage Error) để đánh giá hiệu suất mô hình trên tập kiểm tra.
* **Đánh giá tính ổn định**: Chạy nhiều lần với các tập dữ liệu khác nhau để đảm bảo tính ổn định của mô hình.

**3. Làm sao biết mô hình sẵn sàng để dự báo: ( file Test-LSTM)**

* **Đạt tiêu chí hiệu suất**: Mô hình cần có giá trị RMSE và MAPE đủ thấp trên tập kiểm tra (thường là < 5% cho MAPE được coi là tốt).
* **Đường dự báo hợp lý**: Kiểm tra đồ thị của dữ liệu thực tế so với dự báo để đảm bảo rằng mô hình không quá khớp (overfitting) hoặc chưa khớp (underfitting).
* **Phân tích dư thừa (Residual Analysis)**: Dùng phân tích dư thừa để kiểm tra xem liệu các sai số có phân phối ngẫu nhiên và không còn mô hình nào khác có thể dự đoán được sai số đó.

**4. Các bước chuẩn bị trước khi dự báo:**

* **Tối ưu siêu tham số**: Sử dụng công cụ như Optuna để tối ưu hóa các siêu tham số như số lượng units LSTM, dropout rate, learning rate, số epochs, và batch size.

**Dùng file Run để làm 2 bước cuối**

* **Huấn luyện cuối cùng**: Huấn luyện lại mô hình với toàn bộ tập dữ liệu (kết hợp tập huấn luyện và kiểm tra) với các siêu tham số tối ưu đã tìm được.
* **Dự đoán thử**: Thực hiện dự đoán trên một đoạn dữ liệu có sẵn để xác nhận mô hình hoạt động đúng như mong đợi trước khi dự báo đoạn không có dữ liệu.  
  cái này chưa có trong code, nó nằm sau khi huấn luyện và trước khi dự đoán đoạn 90 ngày

**Quy trình chuẩn hóa cải tiến cho dự báo chuỗi thời gian bằng LSTM**

**1. Xác định mục tiêu và hiểu dữ liệu**

* **Mục tiêu rõ ràng**:
  + Dự báo giá cổ phiếu Tesla trong 90 ngày tới với độ chính xác cao.
  + Đánh giá mô hình dựa trên các chỉ số chính xác như RMSE, MAE, MAPE, hoặc các chỉ số rủi ro.
* **Phân tích dữ liệu**:
  + Kiểm tra tính chất của dữ liệu, bao gồm:
    - **Tính dừng (stationarity)**: Nếu chuỗi không dừng (xu hướng hoặc mùa vụ), cần thực hiện biến đổi như lấy log, sai phân (differencing) trước khi đưa vào mô hình.
    - **Sự tự tương quan (autocorrelation)**: Xác định các mối quan hệ quan trọng trong dữ liệu, sử dụng biểu đồ ACF (Autocorrelation Function) và PACF (Partial ACF).

**2. Tiền xử lý dữ liệu**

* **Xử lý dữ liệu đầu vào**:
  + Chuẩn hóa dữ liệu bằng MinMaxScaler (hoặc StandardScaler nếu phù hợp hơn).
  + Tách dữ liệu thành các chuỗi thời gian (time step) một cách nhất quán.
  + Xử lý các giá trị bị thiếu (nếu có) hoặc ngoại lai ảnh hưởng đến chất lượng dự báo.
* **Chia tập dữ liệu**:
  + **Tập huấn luyện**: 80% dữ liệu ban đầu để mô hình học.
  + **Tập kiểm tra**: 20% còn lại để đánh giá mô hình.
  + **Tập kiểm tra trước tương lai (future validation)**: Một khoảng thời gian nhỏ cuối cùng để đảm bảo rằng mô hình sẵn sàng cho dự báo ngoài thực tế.

**3. Xây dựng mô hình LSTM**

* **Cấu trúc mô hình**:
  + Kiểm tra và so sánh các mô hình khác nhau (1-layer LSTM, stacked LSTM, Bi-directional LSTM).
  + Kết hợp Dropout để tránh overfitting.
  + Điều chỉnh hyperparameter (số lượng units, batch size, learning rate, số epochs) bằng Optuna
* **Đánh giá trên tập kiểm tra**:
  + Dùng các chỉ số như RMSE, MAE, MAPE để đánh giá.
  + Kiểm tra lỗi phân phối bằng cách phân tích phần dư (residual analysis) và đảm bảo không có tự tương quan hoặc xu hướng trong lỗi.

**4. Kiểm thử mô hình**

* **Kiểm thử khả năng dự báo trong ngắn hạn và dài hạn**:
  + Phân tích hiệu năng trong các khung thời gian khác nhau (ví dụ: 15 ngày, 30 ngày, 90 ngày).
  + Đảm bảo rằng mô hình hoạt động tốt không chỉ ở ngắn hạn mà còn dài hạn.
* **Kiểm thử ổn định**:
  + Thực hiện nhiều lần chạy mô hình (trên cùng tập dữ liệu nhưng khác seed ngẫu nhiên) để đảm bảo mô hình ổn định và không phụ thuộc quá nhiều vào một lần huấn luyện.
* **So sánh mô hình**:
  + So sánh hiệu năng của LSTM với các mô hình khác như ARIMA, Prophet hoặc GRU để khẳng định rằng LSTM là lựa chọn tốt nhất.

**5. Khi nào mô hình sẵn sàng để dự báo**

Mô hình sẵn sàng khi:

1. **Hiệu năng đạt yêu cầu**:
   * RMSE hoặc MAE nhỏ và chênh lệch giữa tập huấn luyện và kiểm tra thấp (tránh overfitting).
   * Phần dư không có xu hướng rõ ràng (xác định qua phân tích phần dư).
2. **Độ ổn định cao**:
   * Kết quả dự báo nhất quán khi thay đổi cấu hình hoặc seed ngẫu nhiên.
3. **Khả năng khái quát hóa tốt**:
   * Dự báo chính xác trên tập kiểm tra trước tương lai (future validation).

**6. Quy trình dự báo**

1. Lấy dữ liệu 60 ngày gần nhất (time step) từ tập dữ liệu lịch sử.
2. Sử dụng mô hình đã huấn luyện để dự báo giá trong ngày tiếp theo.
3. Thêm giá dự báo vào chuỗi dữ liệu để dự báo cho ngày kế tiếp (quy trình lặp 90 lần).
4. Chuyển đổi kết quả từ dạng chuẩn hóa về giá trị thực tế (inverse scaling).
5. Hiển thị kết quả dự báo và so sánh với dữ liệu thực tế (nếu có).

**7. Các tiêu chí đánh giá dự báo**

* **Kết quả ngắn hạn và dài hạn**:
  + Dự báo chính xác trong 15 ngày đầu và xu hướng giảm dần mức sai số trong các ngày tiếp theo.
* **So sánh với thực tế**:
  + Nếu có dữ liệu thực tế của giai đoạn đã dự báo, kiểm tra lại để đảm bảo rằng dự đoán nằm trong khoảng sai số mong muốn.
* **Độ tin cậy**:
  + Kiểm tra kết quả với các phương pháp mô phỏng hoặc dự báo bổ sung để xác nhận độ tin cậy.