



Trường Đại học Công nghệ - ĐHQGHN

Weather forecasting

Trần Tiến Nam - 22022594



Tóm tắt



Bối cảnh

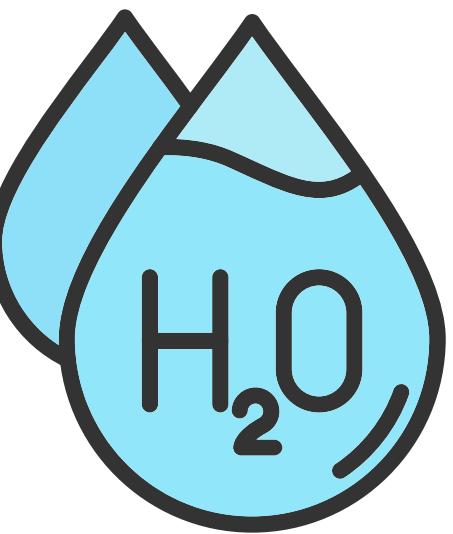
Trong bối cảnh biến đổi khí hậu và nhu cầu lập kế hoạch dựa trên thời tiết (nông nghiệp, du lịch, đời sống hằng ngày), việc dự báo chính xác nhiệt độ và độ ẩm trở nên cực kỳ quan trọng.

Mục tiêu

Xây dựng mô hình dự báo nhiệt độ ($^{\circ}\text{C}$) và độ ẩm (%) trong tương lai dựa trên dữ liệu lịch sử thời tiết, cụ thể:

- Dự báo theo giờ: dự đoán nhiệt độ, độ ẩm từng giờ trong 24-72 giờ tới.
- Dự báo theo ngày: dự đoán nhiệt độ, độ ẩm trung bình/ngày trong 7-14 ngày tới.

Ứng dụng



Quản lý tưới tiêu thông minh

- Tưới ít nhưng hiệu quả - tiết kiệm nước - giảm chi phí
=> Kết hợp cảm biến đất để đưa ra quyết định tưới phù hợp.

Phòng ngừa sâu bệnh theo điều kiện khí hậu

- VD: Nấm mốc thường phát triển khi **độ ẩm** cao và **nhiệt độ** ẩm → Dự báo phum thuốc phòng trước, giảm thiệt hại.

Cảnh báo thời tiết cực đoan

- Mưa lớn → cảnh báo không phun thuốc/làm cỏ → tránh trôi thuốc.
- Lạnh đột ngột → Sưởi ấm cây non hoặc bảo vệ gia súc.
- Gió mạnh → che chắn cho cây trồng vật nuôi, thu hoạch sớm.



Dữ liệu



Input

- Dữ liệu lịch sử gồm:
 - Thời gian (ngày-giờ)
 - Nhiệt độ ($^{\circ}\text{C}$)
 - Độ ẩm (%)
 - Các đặc trưng phụ (optional): áp suất, tốc độ gió, lượng mưa...
- Dữ liệu có thể lấy từ API (ví dụ: Historical Weather API) hoặc file CSV thu thập.

Output

- Với input là chuỗi dữ liệu lịch sử, hệ thống dự đoán:
 - Nhiệt độ từng giờ trong n giờ tới.
 - Độ ẩm từng giờ trong n giờ tới.
 - Nhiệt độ trung bình từng ngày trong m ngày tới.
 - Độ ẩm trung bình từng ngày trong m ngày tới.



Bài toán



Loại bài toán: Time Series Forecasting (Dự báo chuỗi thời gian)

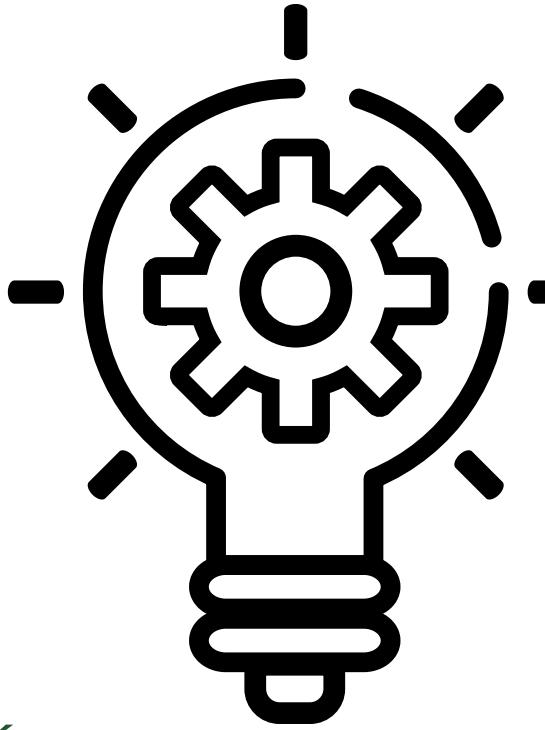
Mục tiêu tối ưu: Giảm sai số giữa giá trị dự đoán và thực tế, đồng thời đánh giá mức độ giải thích dữ liệu của mô hình. Các metric cụ thể:

- MSE (Mean Squared Error)
- MAE (Mean Absolute Error)
- RMSE (Root Mean Squared Error)
- R² (Coefficient of Determination)

Một số mô hình có thể sử dụng

- Classical: ARIMA, SARIMA
- Deep learning: LSTM, GRU, Transformer, Temporal Convolutional Network (TCN)

Phương pháp



Feature Engineering

Tạo đặc trưng thời gian:

- Ngày, tháng, giờ → giúp mô hình hiểu được chu kỳ thời tiết
- Biến nhịp tuần (weekday/weekend), mùa (season)

Tạo đặc trưng trễ (lag features):

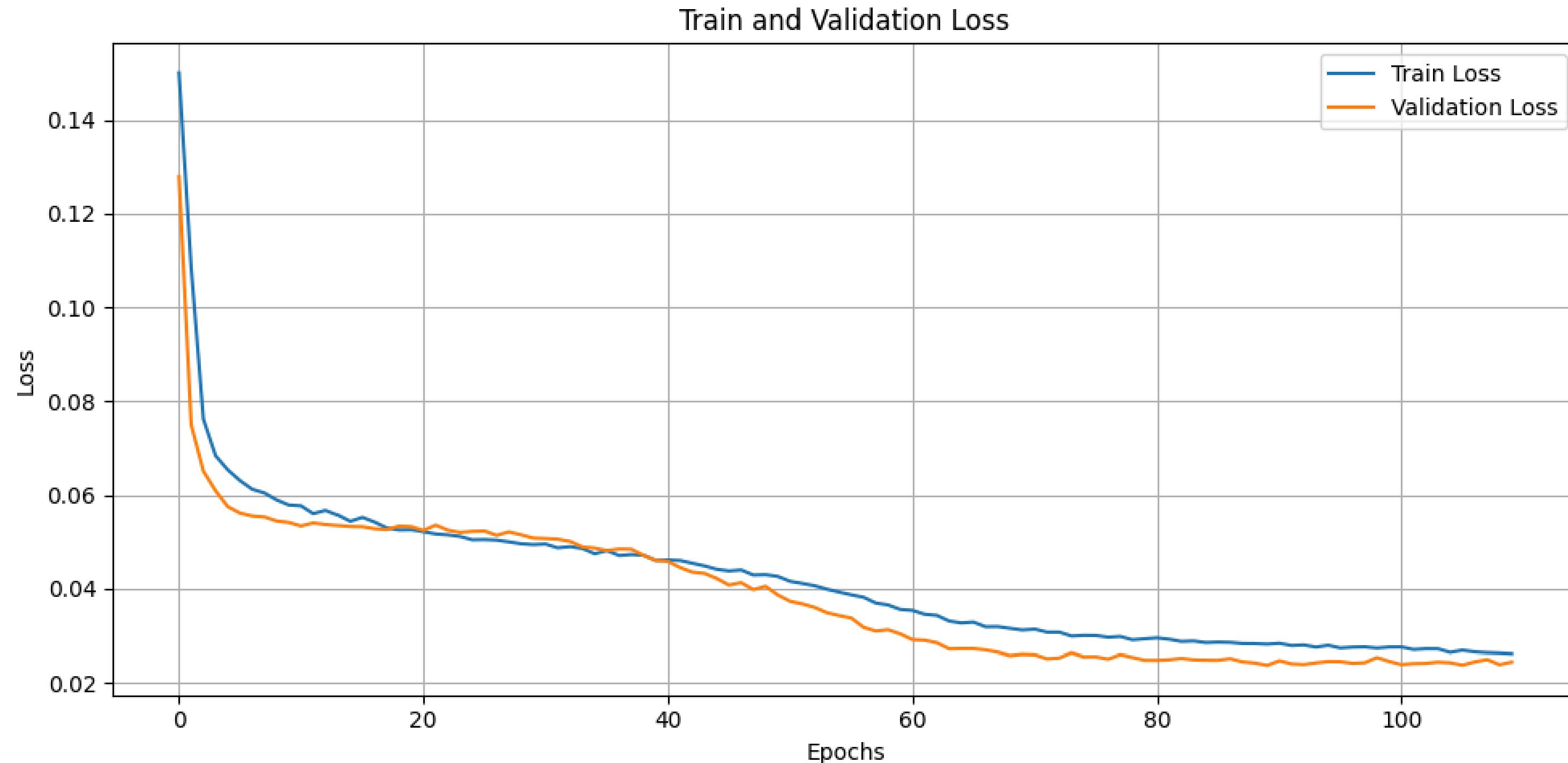
- Dùng nhiệt độ/ độ ẩm của các giờ trước đó (lag 1h, 24h, 1d, 7d)

Tính trung bình trượt (rolling statistics):

- Trung bình trong 24h hoặc 7d gần nhất



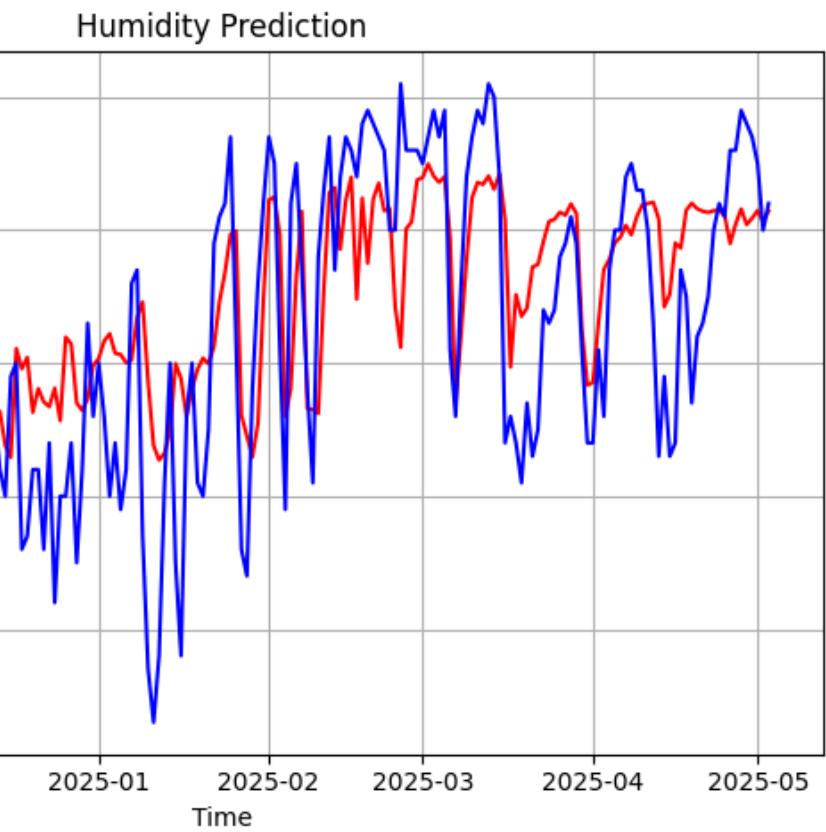
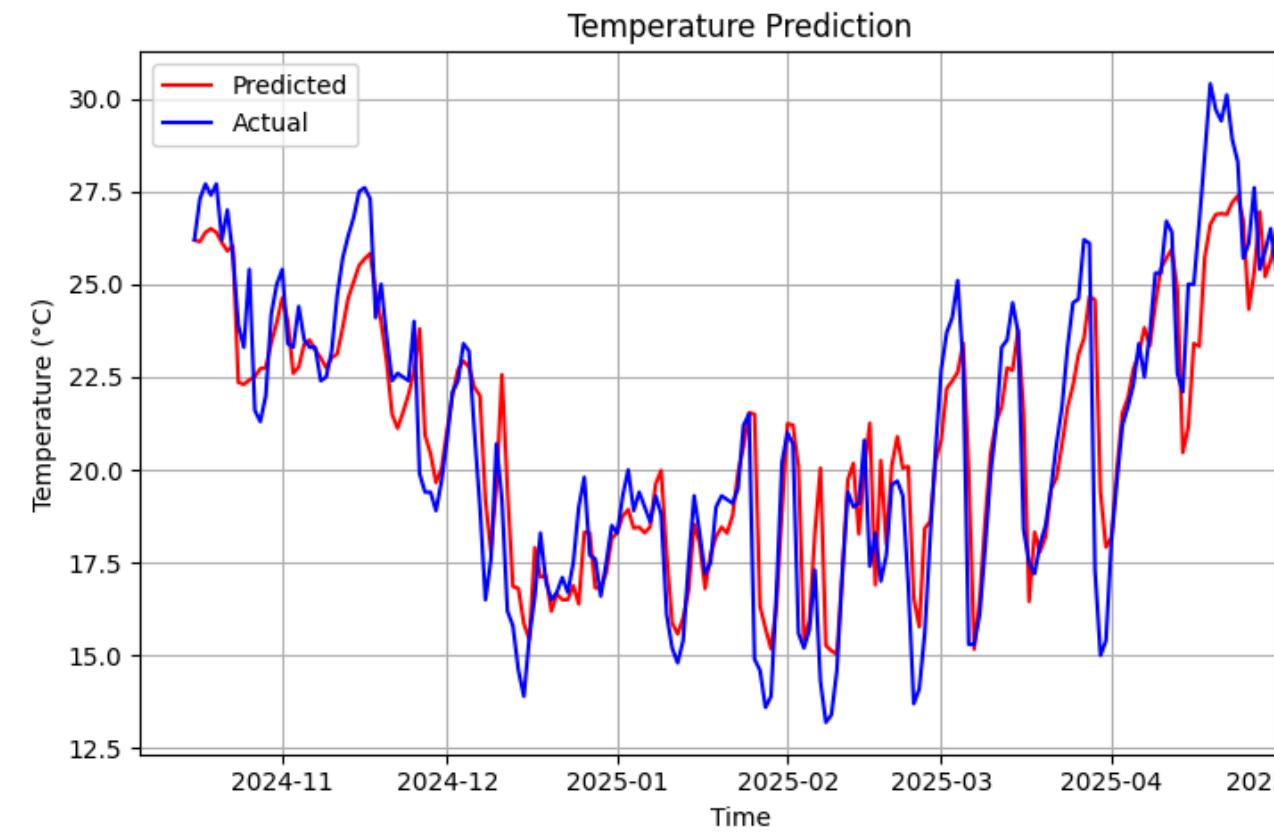
Mô hình LSTM



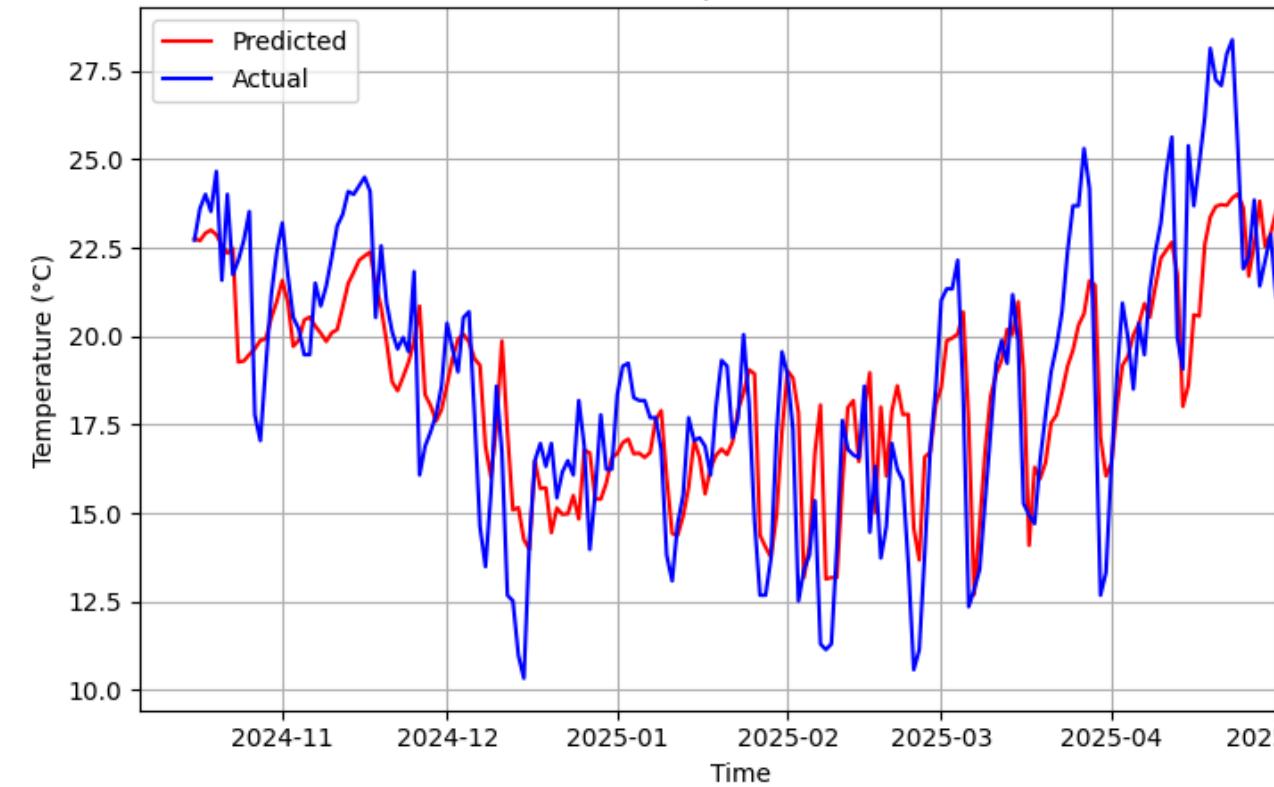
Mô hình LSTM



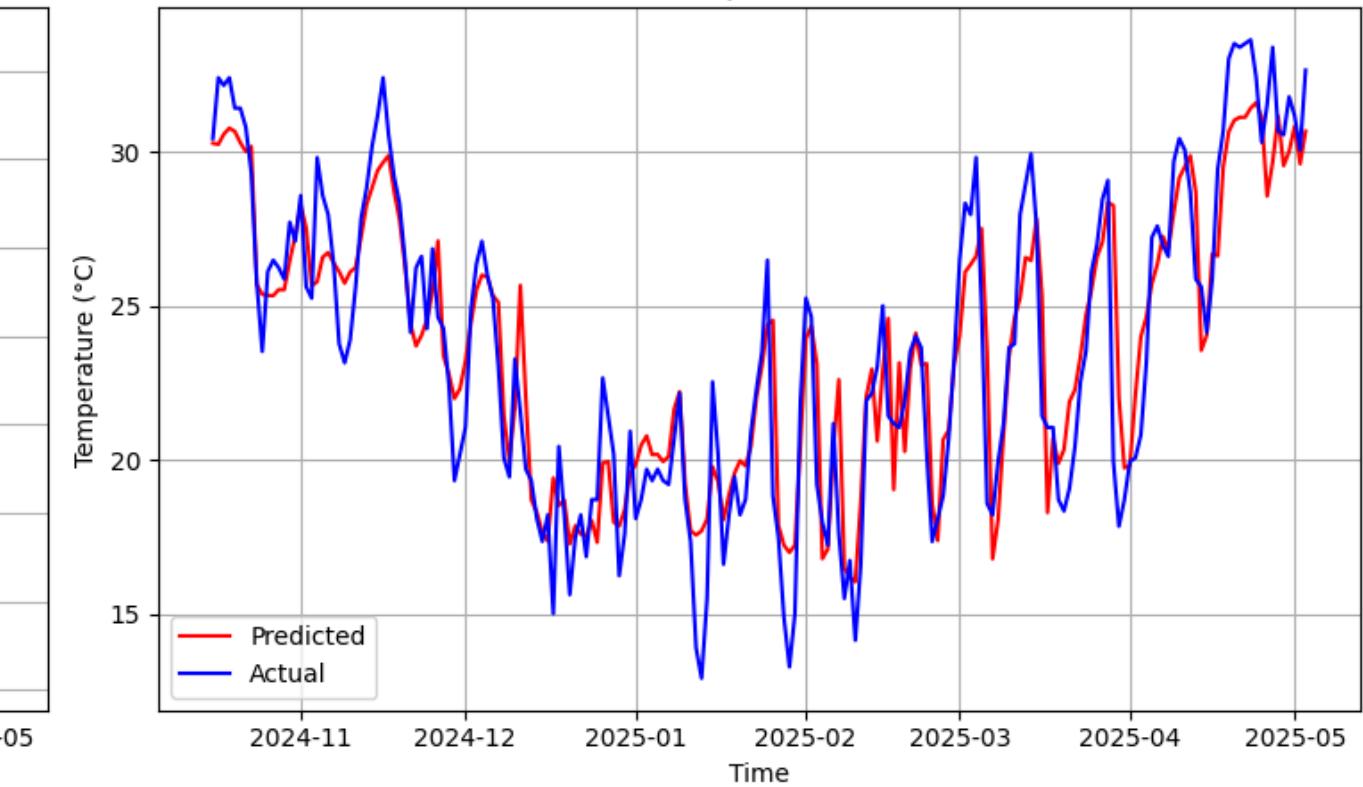
Weather Prediction Results



Minimum Temperature Prediction



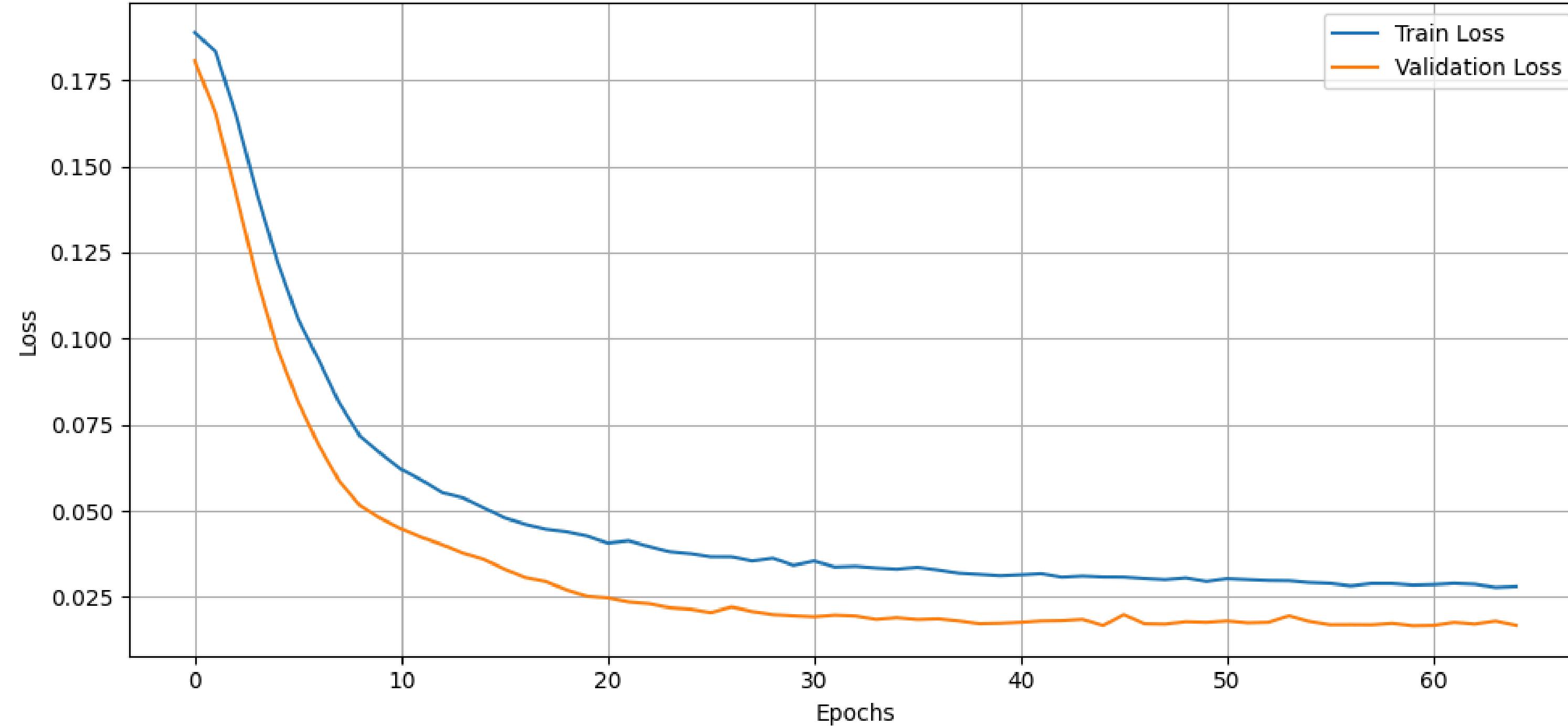
Maximum Temperature Prediction



Mô hình Transformer



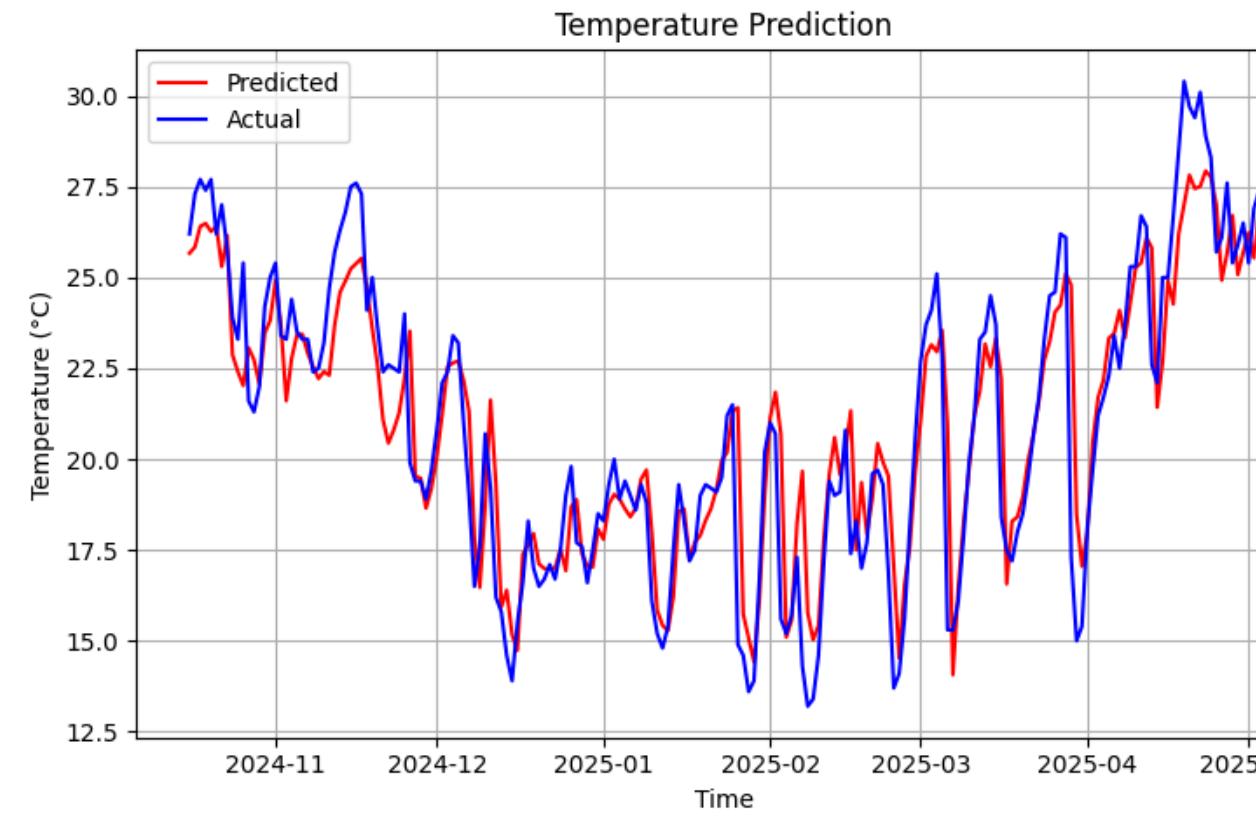
Train and Validation Loss



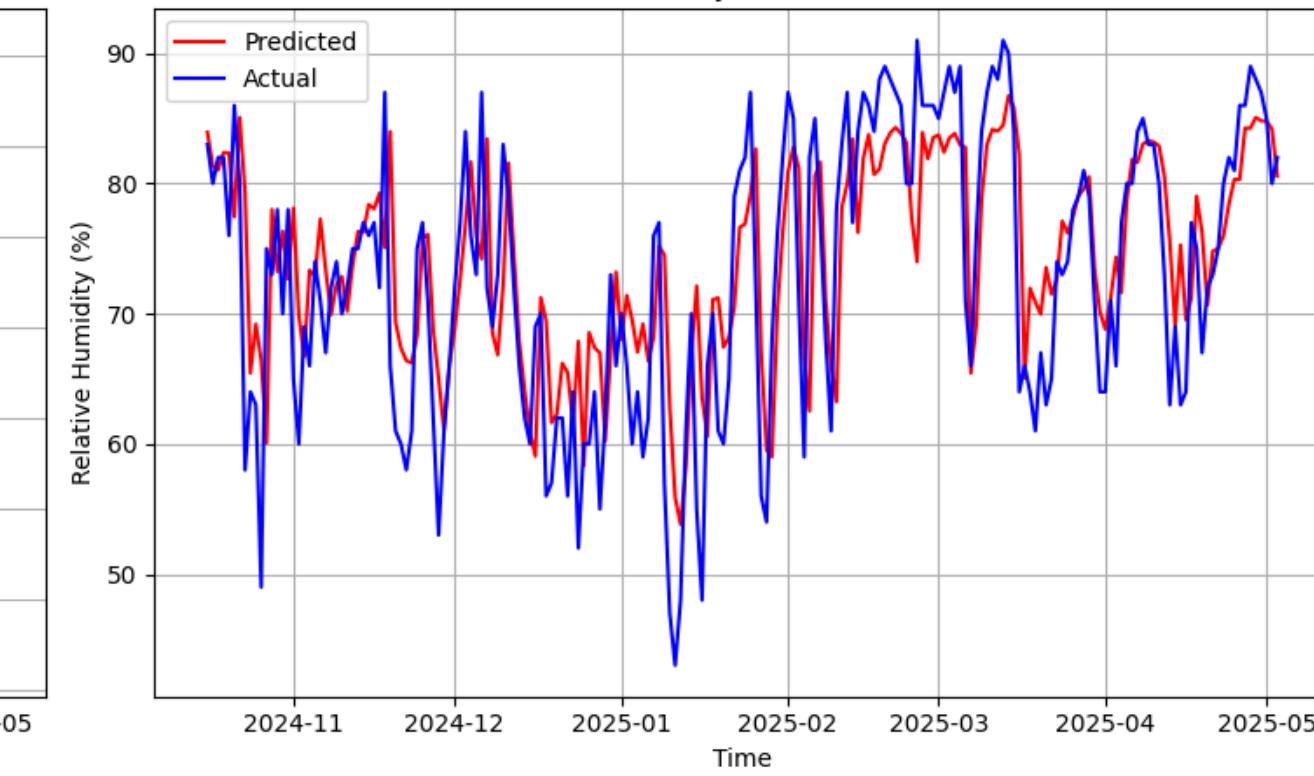
Mô hình Transformer



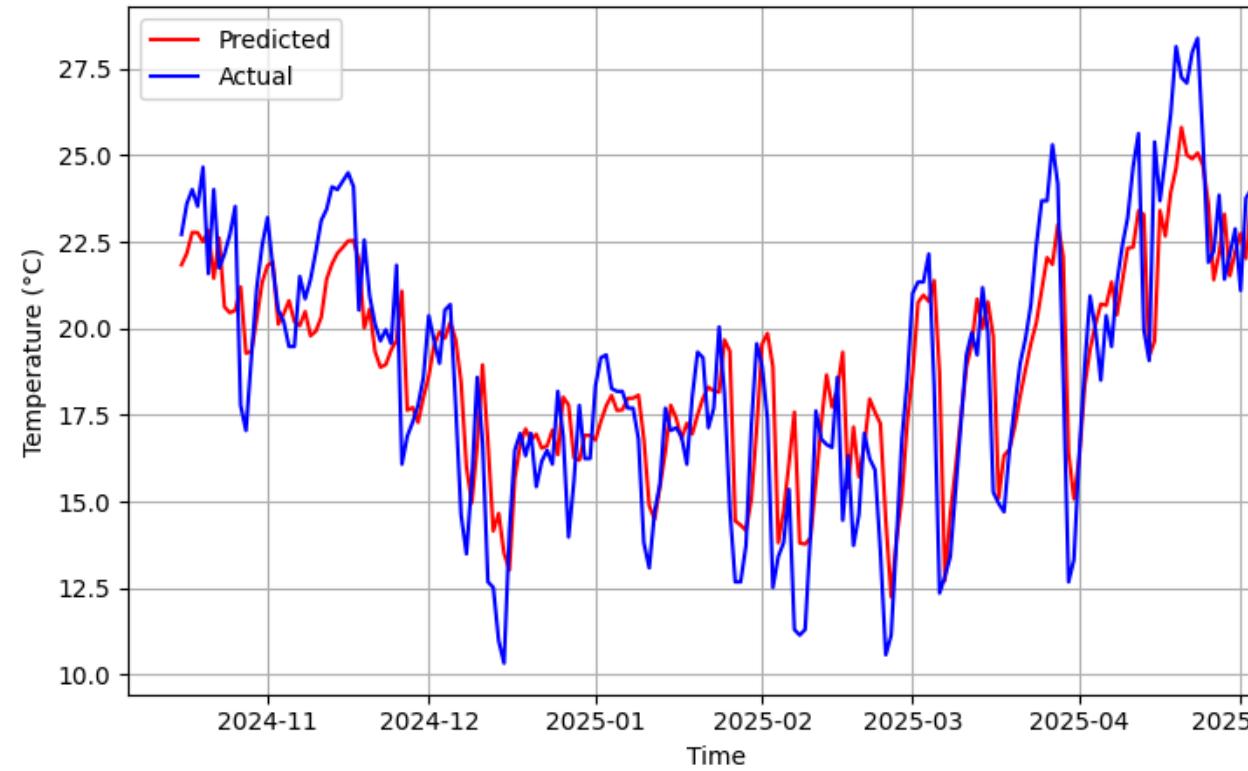
Weather Prediction Results



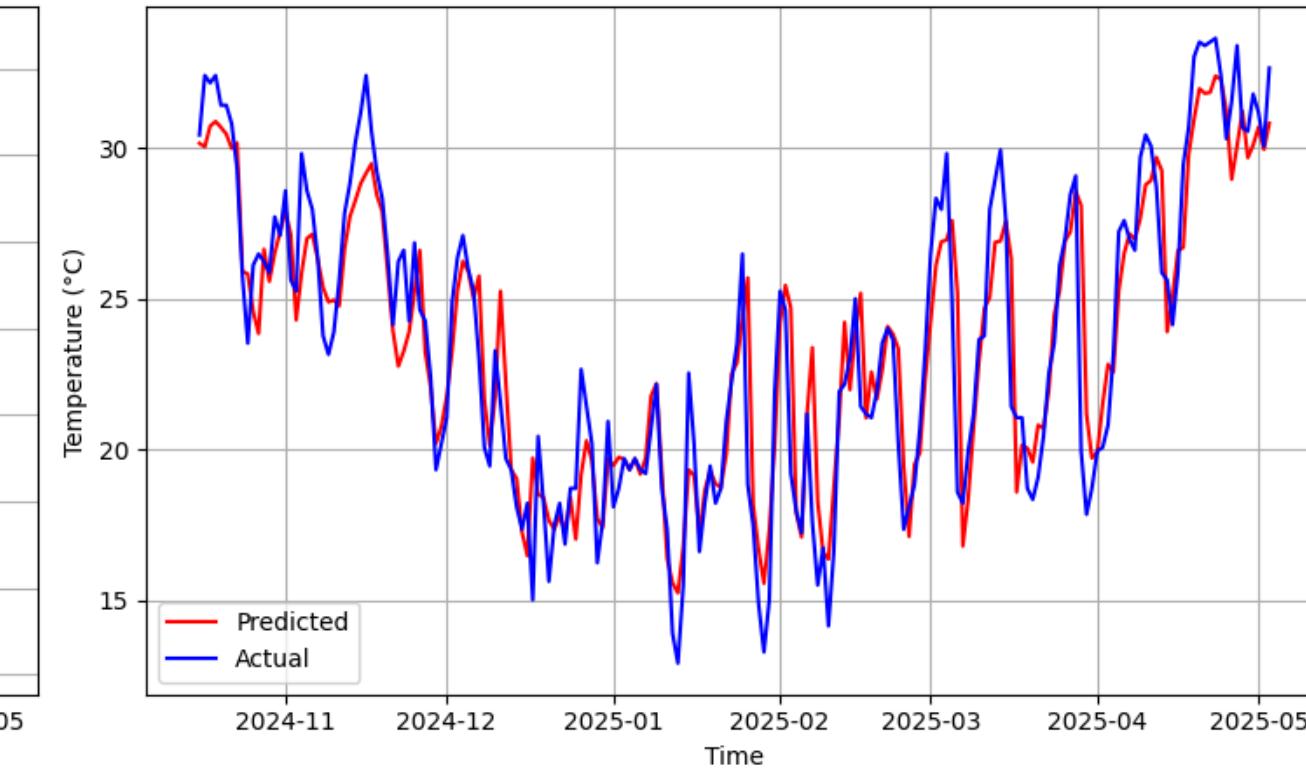
Humidity Prediction



Minimum Temperature Prediction



Maximum Temperature Prediction

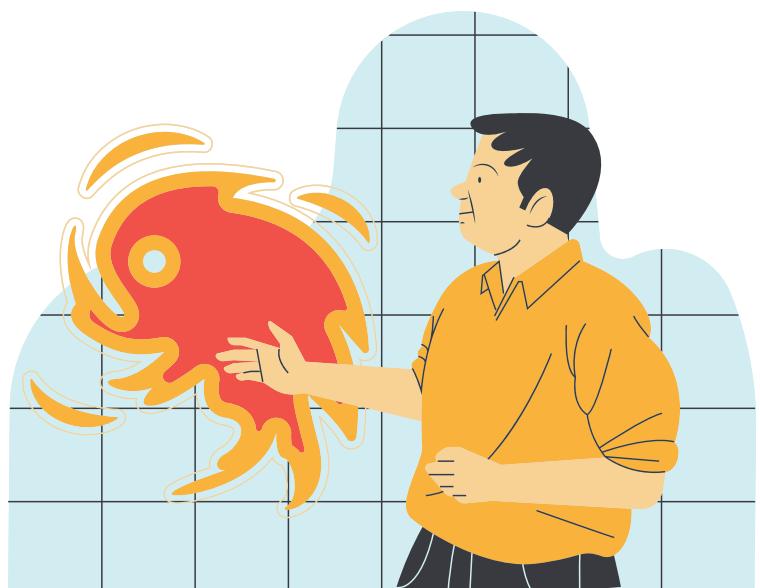


Hàm đánh giá

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$



Predicting daily data with LSTM model. Predicting daily data with Transformer model

Kết quả đánh giá cho mô hình LSTM:

Temperature:
MAE: 1.23 °C
RMSE: 1.62 °C
R²: 0.89

Humidity:
MAE: 5.89 %
RMSE: 7.43 %
R²: 0.49

Min Temperature:
MAE: 1.57 °C
RMSE: 2.00 °C
R²: 0.77

Max Temperature:
MAE: 1.29 °C
RMSE: 1.70 °C
R²: 0.93

Kết quả đánh giá cho mô hình Transformer:

Temperature:
MAE: 1.04 °C
RMSE: 1.42 °C
R²: 0.92

Humidity:
MAE: 4.65 %
RMSE: 6.07 %
R²: 0.66

Min Temperature:
MAE: 1.34 °C
RMSE: 1.75 °C
R²: 0.82

Max Temperature:
MAE: 1.13 °C
RMSE: 1.57 °C
R²: 0.94

Predicting hourly data with LSTM model. Predicting hourly data with Transformer model

Kết quả đánh giá cho mô hình LSTM:

Temperature:
MAE: 0.89 °C
RMSE: 1.15 °C
R²: 0.95

Humidity:
MAE: 2.94 %
RMSE: 3.81 %
R²: 0.95

Kết quả đánh giá cho mô hình Transformer:

Temperature:
MAE: 0.42 °C
RMSE: 0.57 °C
R²: 0.99

Humidity:
MAE: 2.28 %
RMSE: 3.31 %
R²: 0.96



Thách thức



Chất lượng dữ liệu (Data Quality)

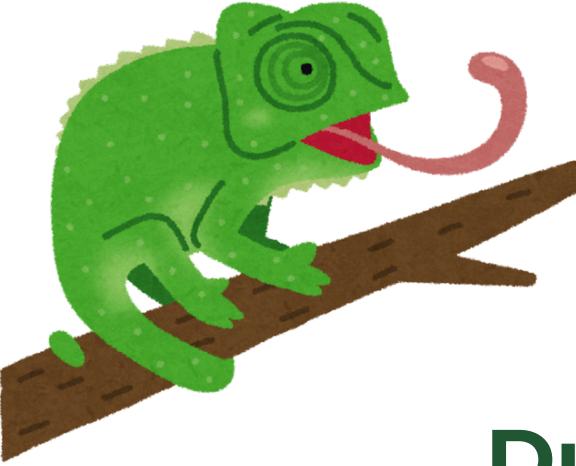
- Dữ liệu lịch sử có thể thiếu sót, nhiễu, không đồng nhất giữa các nguồn.
- Khung thời gian có thể không đều.
- Cần xử lý missing data.

Tính phi tuyến, biến động mạnh (Nonlinearity & Variability)

- Thời tiết là hệ thống rất phi tuyến, chịu tác động của nhiều yếu tố.
- Biến động ngắn hạn rất khó dự báo, nhất là chỉ dựa trên dữ liệu cơ bản như nhiệt độ, độ ẩm.

Overfitting & Generalization (Khả năng tổng quát hóa)

- Mô hình mạnh như LSTM, Transformer dễ fit quá khít dữ liệu train → overfitting.
- Cần kỹ thuật chống overfit: regularization, dropout, early stopping.



Thách thức



Dự báo dài hạn kém chính xác (Long-term Forecasting Challenge)

- Sai số thường tích lũy theo thời gian.

Đặc trưng đầu vào hạn chế (Limited Features)

- Nếu chỉ dùng nhiệt độ & độ ẩm lịch sử, mô hình có thể bị “nhiễu” với các yếu tố quan trọng khác như gió, áp suất, độ che phủ mây, v.v.

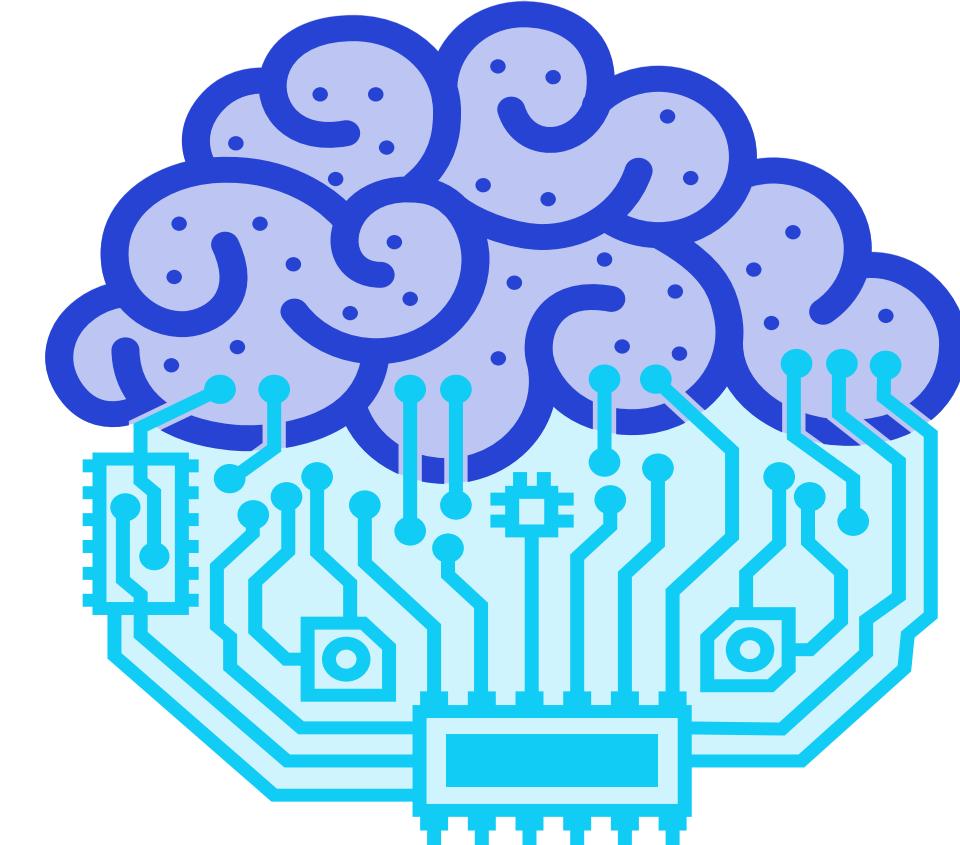
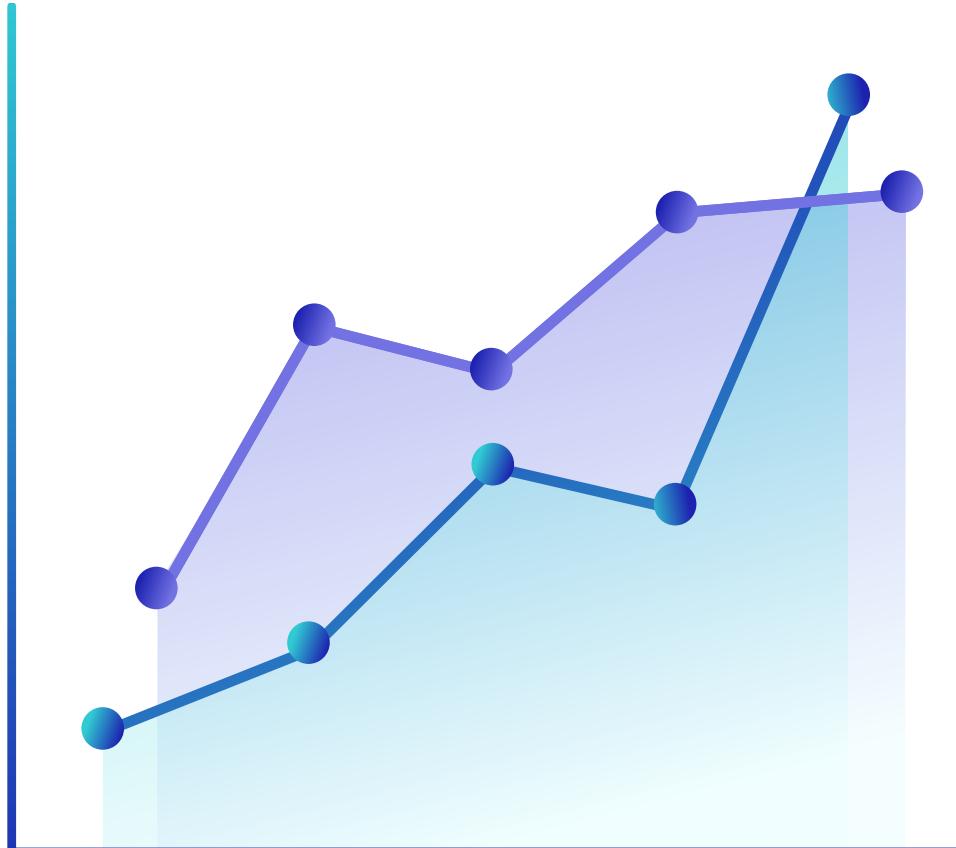
Đánh giá mô hình (Evaluation Challenge)

- Không metric nào hoàn hảo → cần phối hợp nhiều chỉ số (MSE, MAE, RMSE, R²) để đánh giá cân đối.
- Ngoài ra, cần xem biểu đồ thực vs dự đoán để hiểu mô hình sai lệch ở đâu (không chỉ nhìn số).

Kết luận



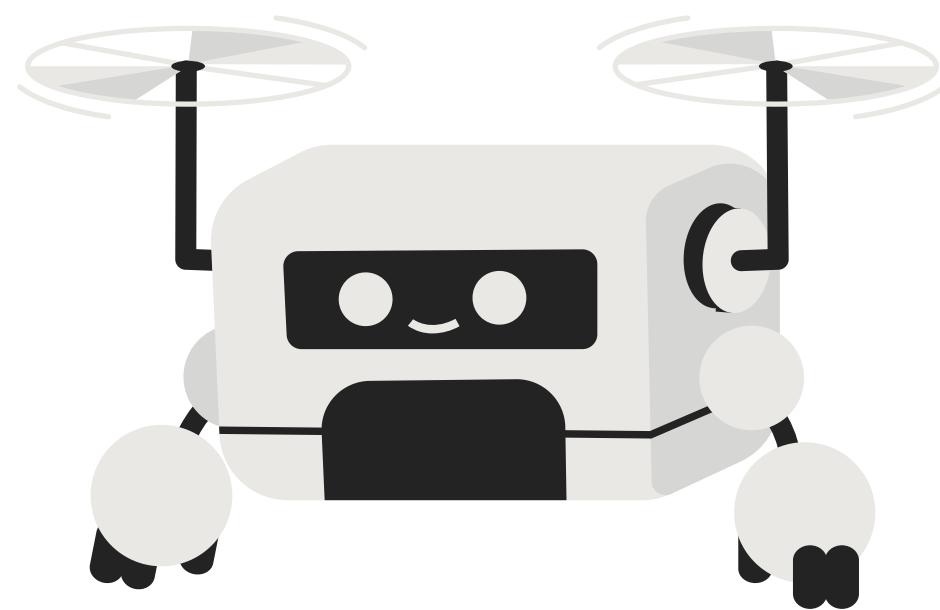
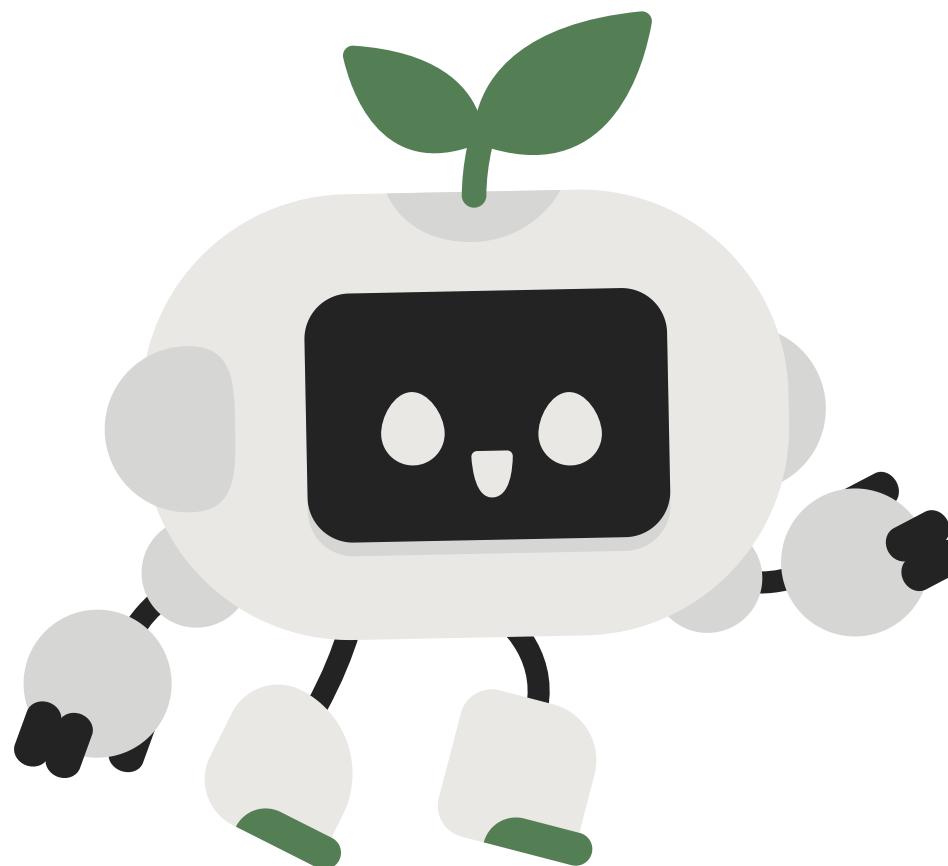
- Các mô hình deep learning (đặc biệt LSTM, Transformer) đã thể hiện hiệu suất vượt trội so với mô hình truyền thống.
- Transformer đạt kết quả cao nhất trên tập dữ liệu hiện tại, nhờ khả năng nắm bắt quan hệ dài hạn trong chuỗi thời gian.
- Tuy nhiên, hiệu quả phụ thuộc mạnh vào chất lượng dữ liệu, chọn đặc trưng, và cách xử lý.



Hướng phát triển



- Bổ sung thêm đặc trưng (gió, áp suất, lượng mưa...)
- Thủ mô hình hybrid & Temporal Fusion Transformer
- Áp dụng học tăng cường để tự tối ưu
- Mở rộng dự báo đa khu vực, cảnh báo thời tiết cực đoan





THANK YOU

