**1. Thư viện và phụ thuộc**

* pandas: Để xử lý dữ liệu.
* numpy: Để thực hiện các phép tính số học.
* torch và torch.nn: Để xây dựng và huấn luyện mô hình Transformer.
* matplotlib.pyplot: Để vẽ biểu đồ kết quả.
* sklearn.preprocessing.StandardScaler: Để chuẩn hóa dữ liệu.

Các thư viện này hỗ trợ việc thao tác dữ liệu, triển khai mô hình và trực quan hóa.

**2. Khai phá dữ liệu**

* **Tập dữ liệu**: Tập dữ liệu (DailyDelhiClimateTrain.csv) chứa dữ liệu khí hậu hàng ngày của Delhi với các cột sau:
  + date: Ngày quan sát (được phân tích dưới dạng datetime64).
  + meantemp: Nhiệt độ trung bình (biến mục tiêu, kiểu float64).
  + humidity: Độ ẩm (float64).
  + wind\_speed: Tốc độ gió (float64).
  + meanpressure: Áp suất khí quyển trung bình (float64).
* Tập dữ liệu có 1.462 bản ghi (không có giá trị thiếu).
* **Trích xuất dữ liệu**:
  + Cột meantemp được trích xuất làm biến mục tiêu (temperature).
  + Chỉ số (dates) được sử dụng cho phân tích thời gian tiềm năng, mặc dù không được sử dụng trực tiếp trong mô hình.

**3. Tiền xử lý dữ liệu**

* **Chuẩn hóa**: Dữ liệu nhiệt độ có khả năng được chuẩn hóa bằng StandardScaler (dựa trên việc sử dụng sklearn.preprocessing.
* StandardScaler và nhãn trục y "Nhiệt độ (chuẩn hóa)" trong biểu đồ).
* **Tạo chuỗi dữ liệu**:
  + Notebook chuẩn bị các cặp đầu vào-đầu ra cho mô hình Transformer:
    - **Đầu vào**: Các chuỗi 30 ngày dữ liệu nhiệt độ (X).
    - **Đầu ra**: Dữ liệu nhiệt độ của 7 ngày tiếp theo (Y).
  + Các chuỗi này được chuyển đổi thành tensor PyTorch (X\_test\_tensor) để đánh giá mô hình.

**4. Định nghĩa mô hình**

* **Mô hình Transformer**:
  + Một mạng nơ-ron dựa trên Transformer được triển khai bằng torch.nn.
  + Kiến trúc mô hình không được hiển thị rõ ràng trong đoạn mã được cung cấp nhưng được ngụ ý là đã định nghĩa ở các cell trước đó (trước phần đánh giá).
  + Mô hình nhận chuỗi nhiệt độ 30 ngày làm đầu vào và dự đoán nhiệt độ cho 7 ngày tiếp theo.
* **Thiết bị**: Mô hình và dữ liệu đầu vào được chuyển sang thiết bị (có thể là GPU nếu có, nếu không là CPU) bằng .to(device).

**5. Đánh giá và dự đoán mô hình**

* Mô hình được đặt ở chế độ đánh giá (model.eval()).
* Dự đoán được thực hiện trên tập kiểm tra (X\_test\_tensor) bằng mô hình Transformer đã huấn luyện.
* Các dự đoán (preds) và giá trị thực (Y\_true) được trích xuất để so sánh.
* Cả preds và Y\_true đều là mảng NumPy sau khi được tách khỏi đồ thị tính toán (.detach().cpu().numpy()).

**6. Trực quan hóa**

* Một biểu đồ được tạo để so sánh nhiệt độ thực và nhiệt độ dự đoán:
  + **Kích thước biểu đồ**: 12x5 inch.
  + **Chi tiết biểu đồ**:
    - Nhiệt độ thực (Y\_true[0]) được vẽ bằng đường liền và điểm đánh dấu hình tròn (marker='o'), được gắn nhãn "Thật".
    - Nhiệt độ dự đoán (preds[0]) được vẽ bằng đường nét đứt và điểm đánh dấu hình chữ 'x' (marker='x'), được gắn nhãn "Dự báo".
  + **Tiêu đề**: "Dự báo 7 ngày nhiệt độ từ 30 ngày đầu vào".
  + **Trục**:
    - Trục X: Gắn nhãn "Ngày", đại diện cho 7 ngày dự đoán.
    - Trục Y: Gắn nhãn "Nhiệt độ (chuẩn hóa)", cho thấy dữ liệu đã được chuẩn hóa.
  + **Lưới và chú thích**: Lưới được thêm để dễ đọc, và chú thích phân biệt giữa giá trị thực và dự đoán.
* Biểu đồ được hiển thị bằng plt.show().

**7. Kết quả đầu ra**

* Notebook bao gồm một hình ảnh PNG của biểu đồ được nhúng trong ô đầu ra, cho thấy sự so sánh giữa nhiệt độ thực và dự đoán.
* Biểu đồ trực quan xác nhận hiệu suất của mô hình, với các giá trị dự đoán thường theo sát xu hướng của giá trị thực, mặc dù có một số sai lệch.
* **Dữ liệu**: Tập dữ liệu sạch (không có giá trị thiếu) và tập trung vào các chỉ số khí hậu hàng ngày, với meantemp là mục tiêu chính.
* **Mô hình**: Mô hình Transformer phù hợp cho dự báo chuỗi thời gian nhờ khả năng nắm bắt các phụ thuộc dài hạn trong dữ liệu chuỗi.
* **Tiền xử lý**: Chuẩn hóa và tạo chuỗi là các bước quan trọng để chuẩn bị dữ liệu cho mô hình.
* **Trực quan hóa**: Biểu đồ truyền tải hiệu quả hiệu suất dự báo của mô hình, với nhãn và điểm đánh dấu rõ ràng để phân biệt giá trị thực và dự đoán.
* **Hạn chế**:
  + Đoạn mã được cung cấp không bao gồm kiến trúc mô hình, vòng lặp huấn luyện hoặc chi tiết siêu tham số, có thể được định nghĩa ở các cell trước.
  + Chỉ một chuỗi (Y\_true[0] và preds[0]) được vẽ, giới hạn đánh giá ở một ví dụ duy nhất.