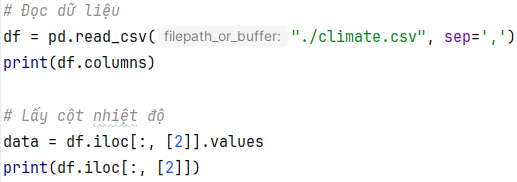
**BÀI TẬP 6**

**XÂY DỰNG MÔ HÌNH HỒI QUY (TH2)**

**--------**

1. **Xây dựng mô hình LSTM cho tập dữ liệu climate.csv (1 cột):**
   1. **Đọc dữ liệu:**



*Hình. Đọc dữ liệu và lấy cột nhiệt độ*

- Dữ liệu được đọc từ file *climate.csv* bao gồm 15 cột như sau:

+ Date Time: Ngày giờ ghi lại dữ liệu.

+ p (mbar): Áp suất không khí trong ngày.

+ T (degC): Nhiệt độ trung bình được đo trong ngày.

+ Tpot (K): Nhiệt độ tiềm năng.

+ Tdew (degC): Điểm sương, nhiệt độ trong không khí bắt đầu ngưng tụ.

+ rh (%): Độ ẩm tương đối.

+ VPmax (mbar): Áp suất hơi bão hòa tối đa.

+ VPact (mbar): Áp suất hơi thực tế.

+ VPdef (mbar): Sự chênh lệch áp suất hơi.

+ sh (g/kg): Độ ẩm tuyệt đối.

+ H2OC (mmol/mol): Nồng độ của hơi nước.

+ rho (g/m\*\*3): Mật độ không khí.

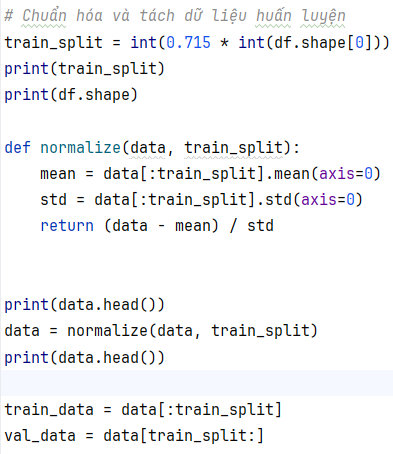
+ wv (m/s): Tốc độ gió trung bình.

+ max. wv (m/s): Tốc độ gió lớn nhất trong ngày.

+ wd (deg): Hướng gió.

- Đối với bài toán này, chỉ lấy dữ liệu ở cột nhiệt độ ‘T (degC)’ để xây dựng mô hình.

* 1. **Chuẩn hóa và tách dữ liệu:**



*Hình. Chuấn hóa và tách dữ liệu*

- Đầu tiên, **train\_split** là số lượng mẫu cho dữ liệu huấn luyện bằng với 71.5% tổng số mẫu dữ liệu **df**.

- Hàm *normalize()* có tác dụng chuẩn hóa dữ liệu, cụ thể:

+ Hàm nhận hai giá trị đầu vào là dữ liệu cần chuẩn hóa **data** và số lượng mẫu huấn luyện **train\_split**.

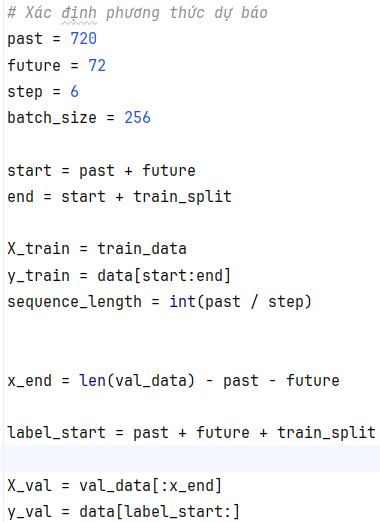
+ **mean** là giá trị trung bình của dữ liệu huấn luyện.

+ **std** là độ lệch chuẩn của dữ liệu huấn luyện.

+ Kết quả của hàm là dữ liệu đã được chuẩn hóa bằng cách trừ trung bình và chia cho độ lệch chuẩn.

- Thực hiện tách dữ liệu theo tỉ lệ **train\_split** với 71.5% cho **train\_data** và 28.5% cho **val\_data**.

* 1. **Xác định phương thức dự đoán:**



*Hình. Xác định phương thức dự đoán*

- **past** là số lượng điểm dữ liệu trong quá khứ mà mô hình sẽ dùng để dự đoán các giá trị tiếp theo. Trong bài này, mô hình sẽ sử dụng 720 điểm dữ liệu tương đương 720 khoảng thời gian với mỗi khoảng là 10 phút, tức là 5 ngày.

- **future** là số lượng điểm dữ liệu trong tương lai mà chúng ta muốn dự đoán. Với bài này, mô hình sẽ dự đoán 12 giờ tiếp theo.

- **step** là khoảng cách giữa các điểm dữ liệu trong chuỗi đầu vào.

- **batch\_size** là kích thước của mỗi batch trong quá trình huấn luyện.

- **start** và **end** lần lượt là vị trí bắt đầu và kết thúc cho **y\_train**.

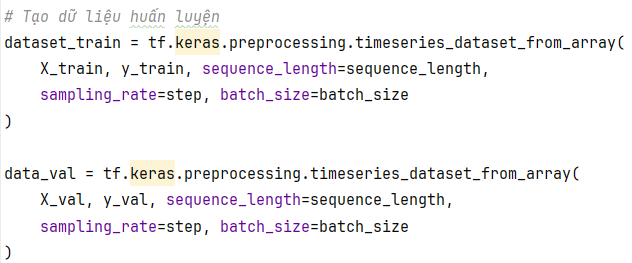
- **X\_train** sẽ lấy toàn bộ **train\_data** làm dữ liệu đầu vào. **y\_train** nhận dữ liệu từ vị trí start đến end của dữ liệu **data**.

- **sequence\_length** là độ dài của chuỗi thời gian.

- **x\_end** là giới hạn cuối của **X\_val**. **Label\_start** là vị trí bắt đầu cho **y\_val**.

- **X\_val** lấy dữ liệu từ đầu đến **x\_end** của **val\_data**. **y\_val** lấy dữ liệu từ **label\_start** đến hết của dữ liệu **data**.

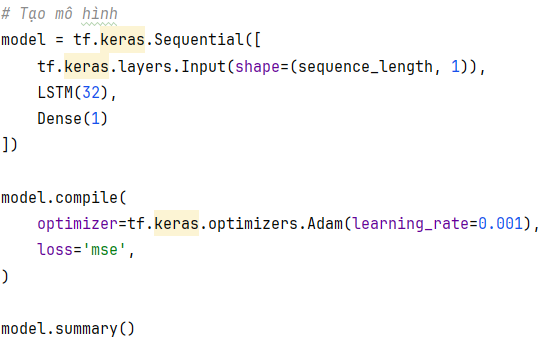
* 1. **Tạo dữ liệu huấn luyện:**



*Hình. Tạo dữ liệu huấn luyện*

Tạo dataset chuỗi thời gian cho **X\_train** và **y\_train** cho **dataset\_train**, **X\_val** và **y\_val** cho **data\_val**.

* 1. **Tạo mô hình:**



*Hình. Tạo mô hình*

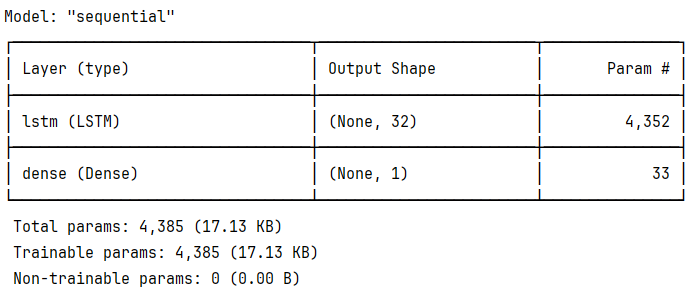
**-** Sử dụng lớp **Sequential()** để xây dựng mô hình:

+ Tầng **Input** xác định kích thước của chuỗi đầu vào cho mô hình với **sequence\_length** là số lượng điểm dữ liệu trong mỗi chuỗi đầu vào và ‘1’ là số lượng đặc trưng cho mỗi điểm dữ liệu do bài này chỉ nhận một đặc trưng là nhiệt độ.

+ Tầng **LSTM** có 32 units, còn được gọi là số nơ-ron.

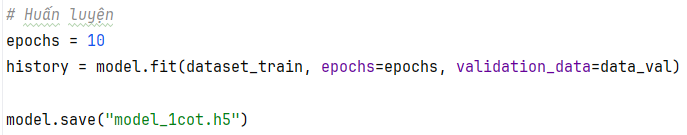
+ Tầng đầu ra **Dense** sẽ lấy đầu ra từ tầng **LSTM** và chuyển thành giá trị dự báo mong muốn.

- Biên dịch mô hình với optimizer Adam và hàm lỗi MSE.



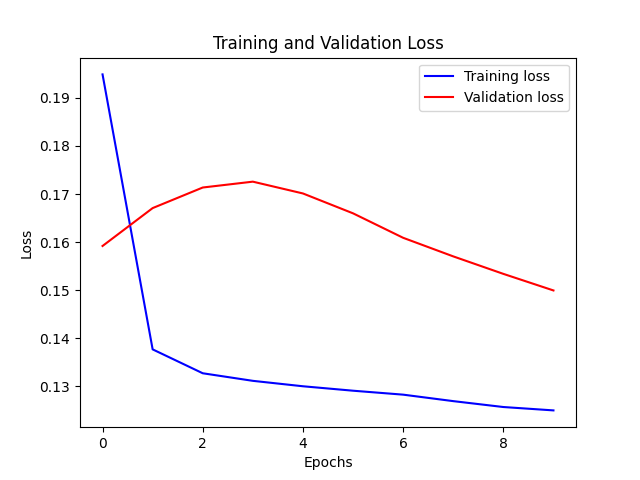
*Hình. Kiến trúc mô hình*

* 1. **Huấn luyện mô hình:**



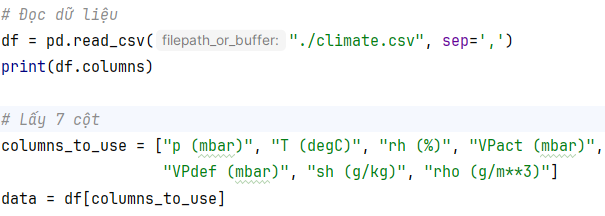
*Hình. Huấn luyện mô hình*

Mô hình sẽ được huấn luyện qua 10 epochs với validata\_data là tập data\_val và được lưu với tên “model\_1cot.h5”.



*Hình. Kết quả huấn luyện*

1. **Xây dựng mô hình LSTM cho tập dữ liệu climate.csv (7 cột):**
   1. **Đọc dữ liệu:**

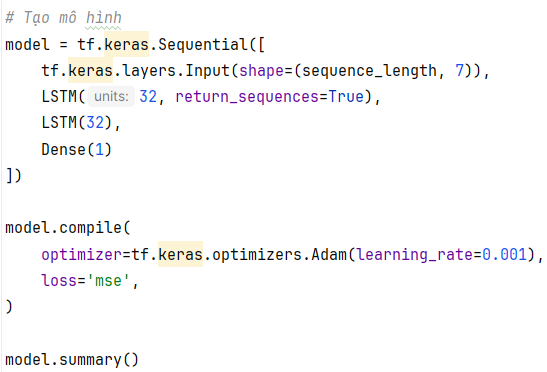


*Hình. Đọc dữ liệu và tạo từ điển*

- Dữ liệu được đọc qua file csv “climate.csv” với 15 cột như mô tả như bài toán trước.

- Đối với bài toán này, những cột được đem sử dụng để xây dựng mô hình bao gồm các cột như hình trên.

* 1. **Xây dựng mô hình:**



*Hình. Xây dựng mô hình*

**-** Sử dụng lớp **Sequential()** để xây dựng mô hình:

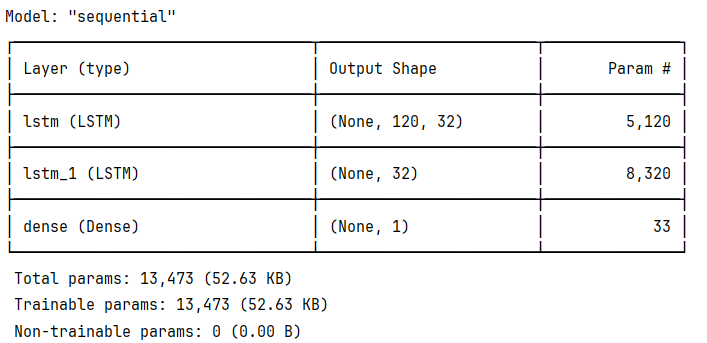
+ Tầng **Input** xác định kích thước của chuỗi đầu vào cho mô hình với **sequence\_length** là số lượng điểm dữ liệu trong mỗi chuỗi đầu vào và ‘7’ là số lượng đặc trưng cho mỗi điểm dữ liệu do bài này nhận vào bảy đặc trưng.

+ Tầng **LSTM** đầu tiên có 32 units, còn được gọi là số nơ-ron. Tham số *return\_sequence* để trả về toàn bộ đầu ra theo từng timestep.

+ Tầng **LSTM** thứ hai cũng có 32 units, nhưng tầng này chỉ chuỗi đầu ra cuối cùng.

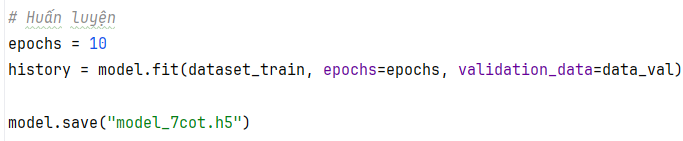
+ Tầng đầu ra **Dense** sẽ lấy đầu ra từ tầng **LSTM** và chuyển thành giá trị dự báo mong muốn.

- Biên dịch mô hình với optimizer Adam và hàm lỗi MSE.



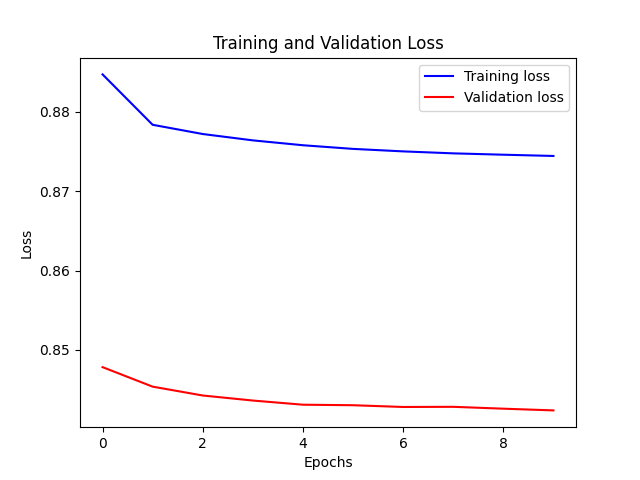
*Hình. Kiến trúc mô hình*

* 1. **Huấn luyện mô hình:**



*Hình. Huấn luyện mô hình*

Mô hình sẽ được huấn luyện qua 10 epochs với validata\_data là tập data\_val và được lưu với tên “model\_7cot.h5”.



*Hình. Kết quả huấn luyện*

**--- HẾT ---**