





Trần Thị Minh Ánh



Thực Tập Chuyên Môn



Type here to



PHÂN TÍCH, XỬ LÝ VÀ DỰ ĐOÁN LƯU LƯỢNG THAM GIA GIAO THÔNG DỰA TRÊN TẬP DỮ LIỆU TRỰC TUYẾN



GV Hướng Dẫn:

Th.S Phạm Thị Dung
K.S Trần Quốc Khánh



Type here to search



7:50 AM
01/07/2023



I. Thuật Toán



III. Kết Quả Xây Dựng Mô Hình

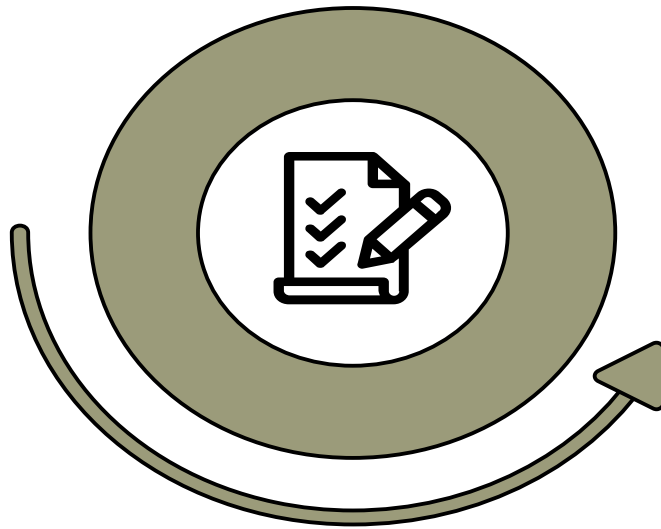


II. Phân Tích Và Xử Lý Dữ Liệu

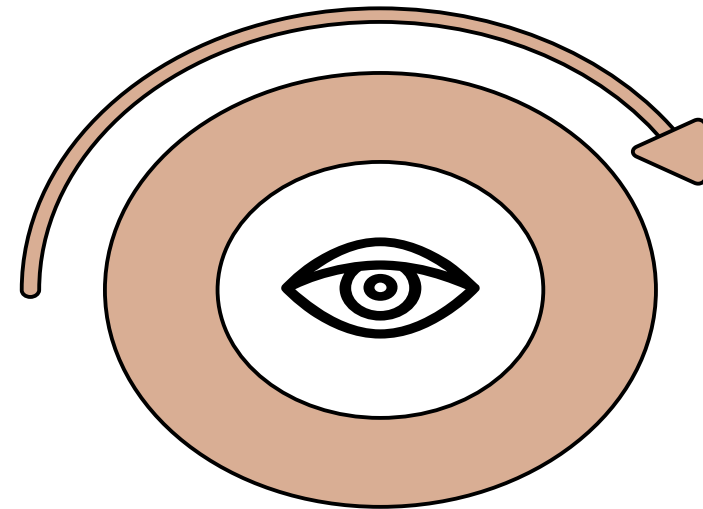


IV. Kết Luận

THUẬT TOÁN



1. Phương Pháp Thống Kê




2. Thuật Toán Backpropagation

THUẬT TOÁN

The icon consists of a green circle with a white document icon inside, which has a checkmark. A curved arrow surrounds the circle, pointing clockwise.

1. Phương Pháp Thống Kê

The icon is a black and white line drawing of a clipboard with a pencil resting on it.

Sử dụng các kỹ thuật thống kê có thể nhận ra các mô hình giao thông ở các tỷ lệ khác nhau, chẳng hạn như trong ngày, vào các ngày trong tuần, theo mùa, vv. So với các phương pháp học máy, chúng thường nhanh hơn, đơn giản và dễ thực hiện hơn. Tuy nhiên, vì chúng không xử lý được nhiều dữ liệu đa biến hơn, nên độ chính xác của chúng thấp hơn.

1. Phương Pháp Thống Kê

DESCRIBE()

COUNT = số lượng giá trị không phải NaN trong cột

MEAN= tổng các giá trị / tổng số mẫu

*ĐỘ LỆCH CHUẨN STD = $\text{sqrt}((1/N) * \text{sum}((x_i - \text{mean})^2))$*

MIN: Giá trị nhỏ nhất trong cột.

MAX: Giá trị lớn nhất trong cột.

Các phân vị 25%, 50% (hoặc median) và 75% của dữ liệu.

THUẬT TOÁN

1. Phương Pháp Thống Kê

	count	mean	std	min	25%	50%	75%	max
temp	48204.0	281.205870	13.338232	0.0	272.16	282.45	291.806	310.07
rain_1h	48204.0	0.334264	44.789133	0.0	0.00	0.00	0.000	9831.30
snow_1h	48204.0	0.000222	0.008168	0.0	0.00	0.00	0.000	0.51
clouds_all	48204.0	49.362231	39.015750	0.0	1.00	64.00	90.000	100.00
traffic_volume	48204.0	3259.818355	1986.860670	0.0	1193.00	3380.00	4933.000	7280.00



2. Thuật Toán Backpropagation

- ✓ Backpropagation (phản hồi ngược) là một thuật toán quan trọng được sử dụng trong huấn luyện mạng nơ-ron để tính toán đạo hàm của hàm mất mát theo các trọng số của mô hình.
- ✓ Thuật toán này cho phép chúng ta điều chỉnh các trọng số để giảm thiểu lỗi và cải thiện hiệu suất của mô hình

THUẬT TOÁN

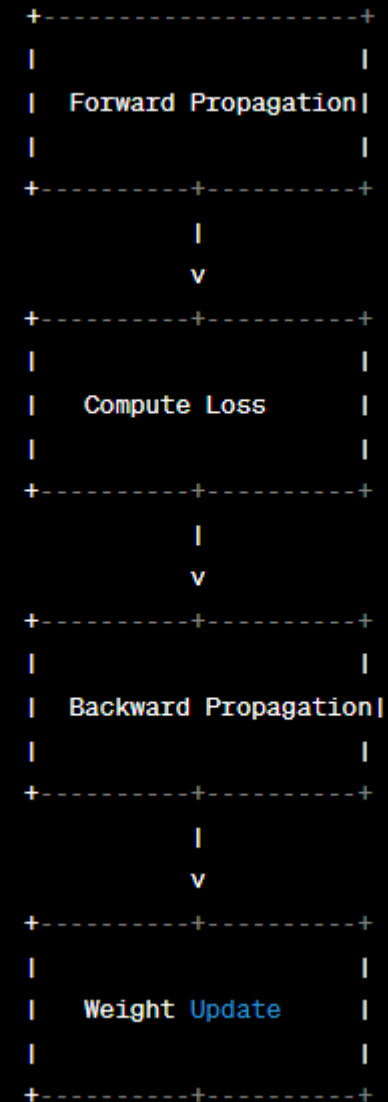


2. Thuật Toán Backpropagation

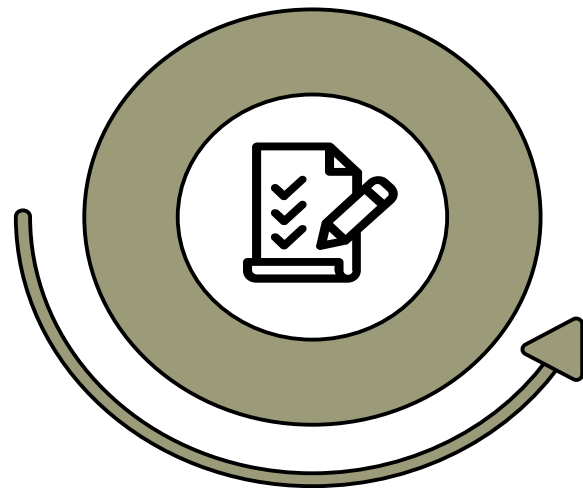
Trong sơ đồ trên, quá trình áp dụng thuật toán backpropagation trong mô hình LSTM được biểu diễn như sau:

1. Forward Propagation: Đầu vào được truyền qua mạng LSTM để tính toán đầu ra dự đoán.
2. Compute Loss: Đầu ra dự đoán được so sánh với đầu ra thực tế để tính toán độ lỗi.
3. Backward Propagation: Từ độ lỗi, thuật toán backpropagation tính toán đạo hàm của hàm mất mát theo các trọng số trong mạng LSTM.
4. Weight Update: Các trọng số được cập nhật dựa trên đạo hàm để giảm thiểu độ lỗi.

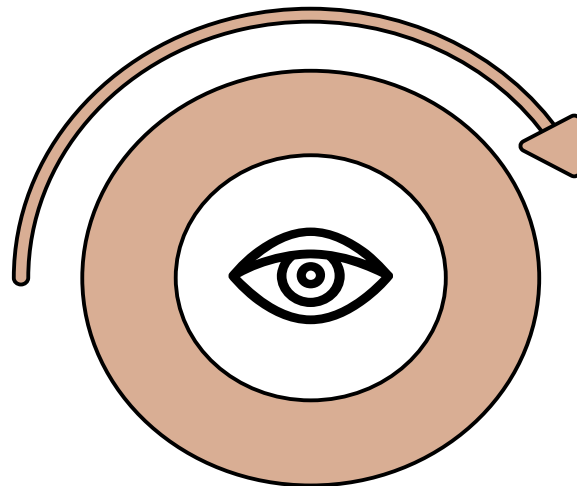
Quá trình Forward Propagation, Compute Loss, Backward Propagation và Weight Update được lặp lại trong quá trình huấn luyện để cải thiện hiệu suất của mô hình LSTM trong việc dự đoán chuỗi thời gian.



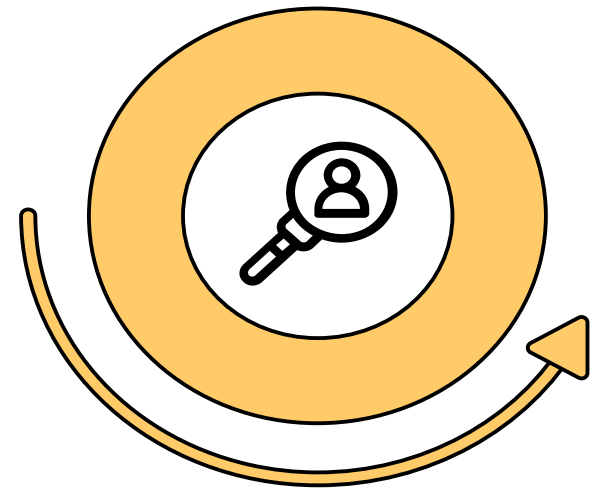
PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU



1. Bộ Dữ Liệu



2. Phân Tích Dữ Liệu



3. Xử Lý Và Làm Sạch Dữ Liệu

PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU



1. Bộ Dữ Liệu

- Bộ dữ liệu “Metro Interstate”.
- Bộ dữ liệu về lưu lượng giao thông liên bang chứa thông tin về lưu lượng giao thông hàng giờ trên làn đường hướng Tây của Xa lộ Liên tiểu bang-94 (I-94) ở Hoa Kỳ.
- Bộ dữ liệu bao gồm các báo cáo thời tiết và nhiệt độ hàng giờ từ năm 2012 đến 2018

PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU



2. Phân Tích Dữ Liệu

Không có dữ liệu bị thiếu, nhưng điều đó không có nghĩa là không có dữ liệu không nhất quán

Có các bản ghi nhiệt độ ở số 0 tuyệt đối, dữ liệu không nhất quán rõ ràng.

Rain_1h và Snow_1h có nhiều số 0 và phân phối của chúng không được xác định rõ ràng, vì trong những lúc cao điểm, chúng có hồ sơ cao.

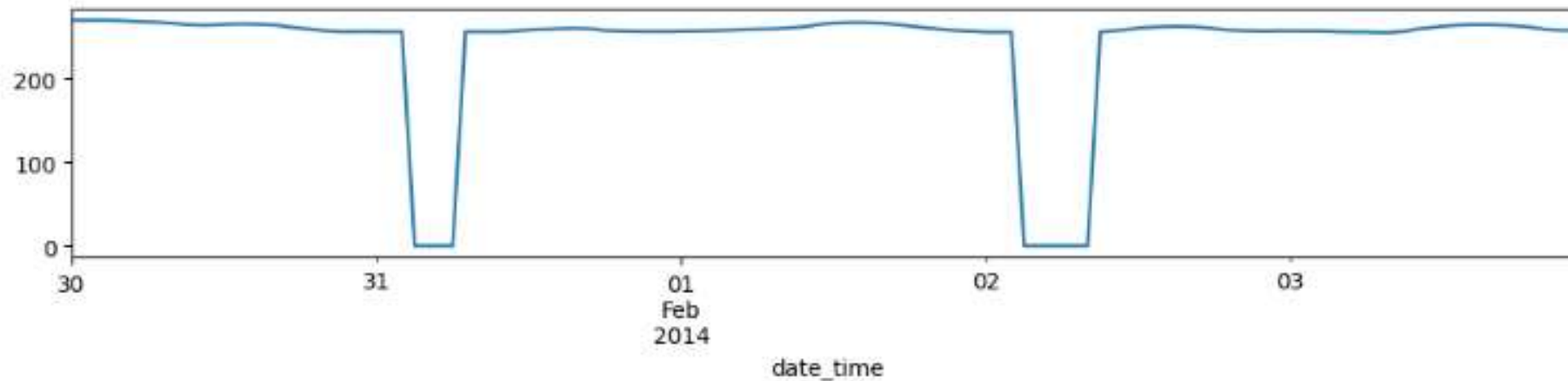
Dữ liệu bị thiếu bản ghi giữa 2013-10 và 2013-11 và dữ liệu bị thiếu bản ghi giữa 2014-08-08 và 2015-06-11

99.9% dữ liệu ngày lễ là None, và dữ liệu khác được truyền bá multiple holidays

PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU



3. Xử Lý Dữ Liệu



Nhiệt Độ

→ 2 ngày có giá trị sai [2014-01-31, 2014-02-02]

→ Các giá trị nhiệt độ bằng 0 là sai số

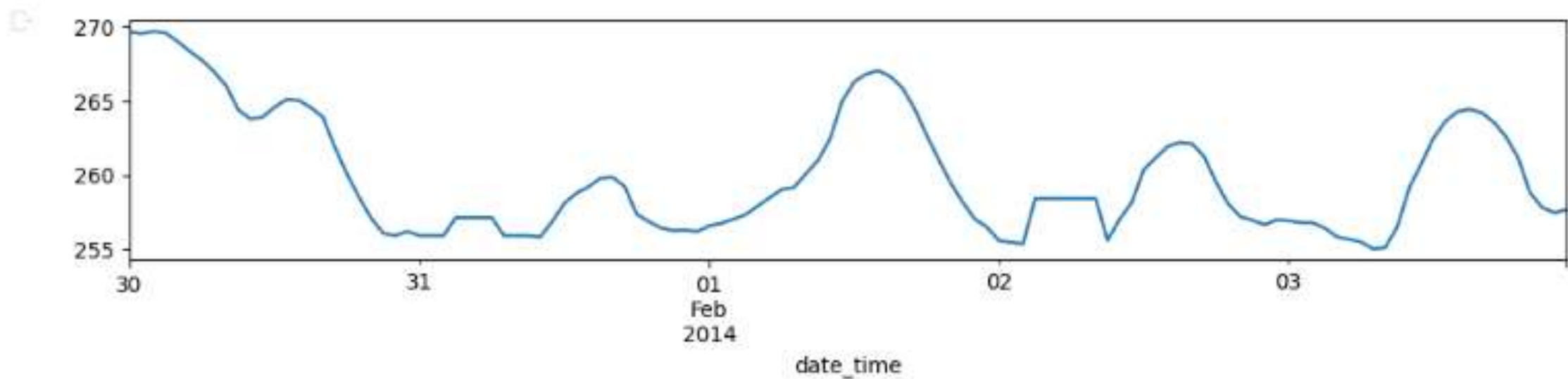
=> Thay thế bằng giá trị trung bình của ngày.

PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU



3. Xử Lý Dữ Liệu

Nhiệt Độ

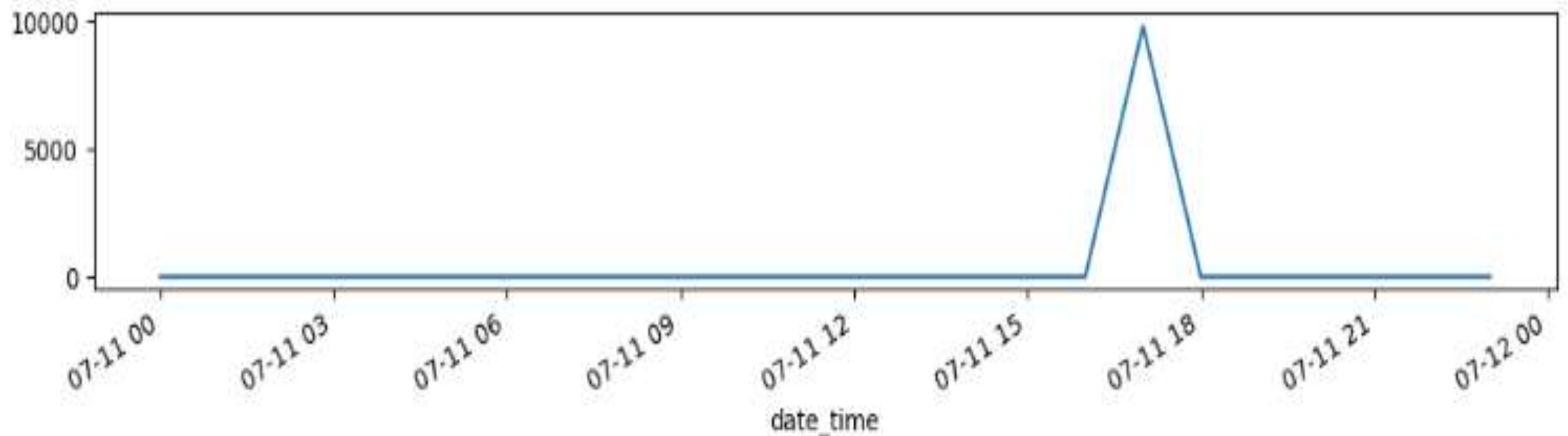


=> Nhiệt độ đã được xử lý ở 2 ngày [2014-01-31, 2014-02-02]

PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU



3. Xử Lý Dữ Liệu



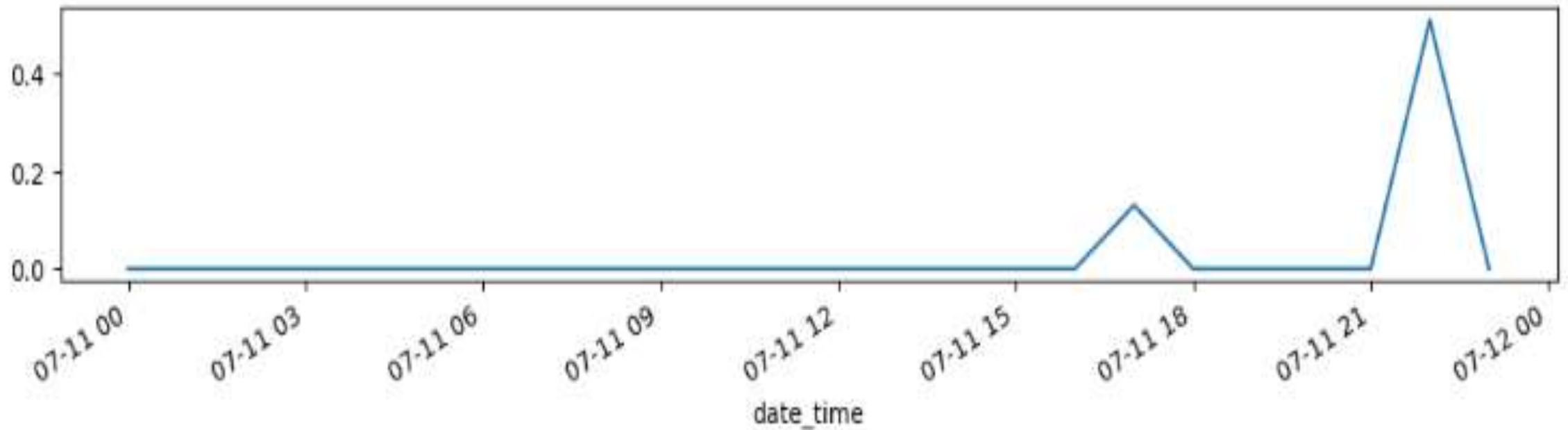
Mưa

Mưa chỉ có 1 bản ghi với giá trị cực trị => Em sẽ chỉ đặt thành giá trị trung bình chung của mưa.

PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU



3. Xử Lý Dữ Liệu

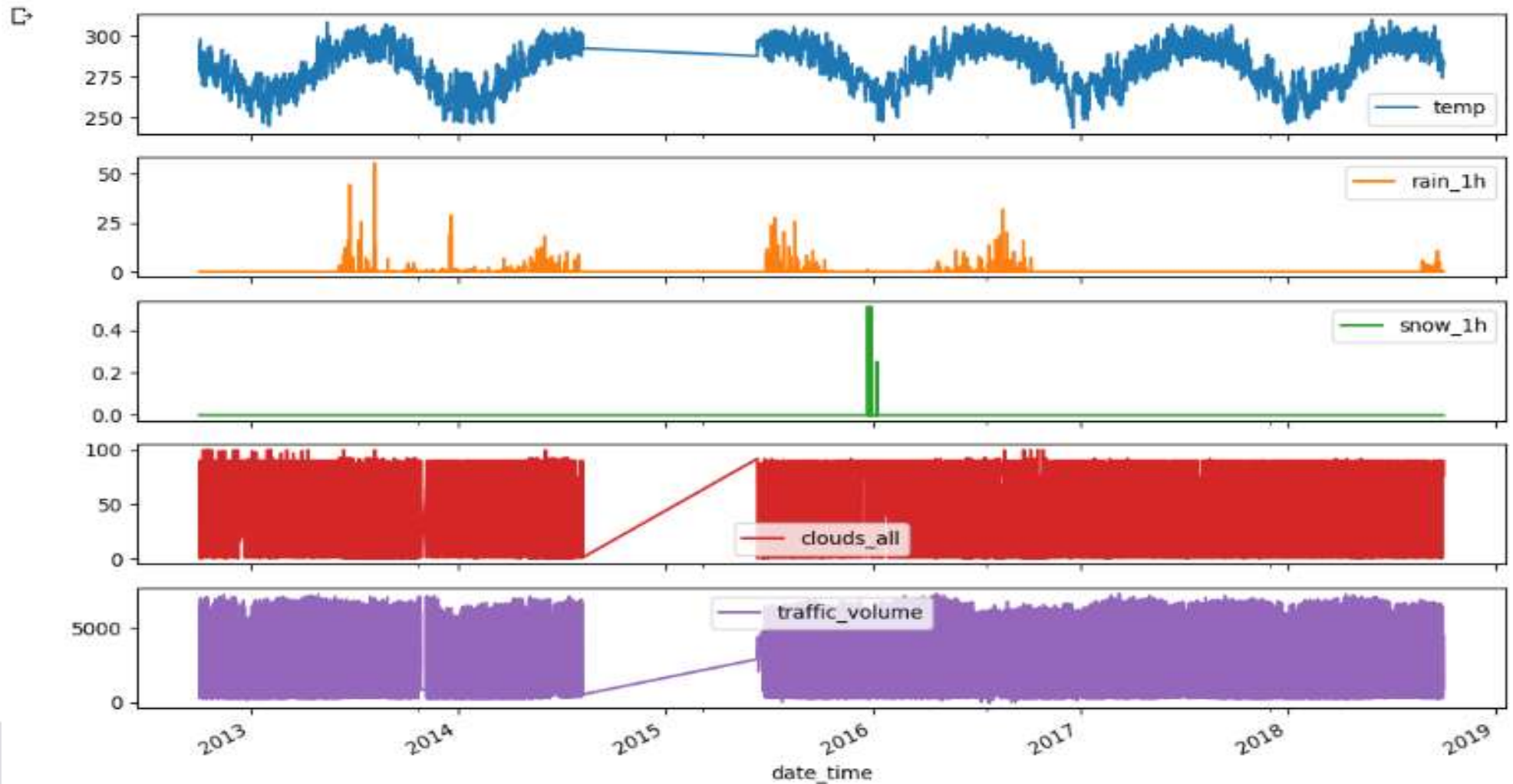


Mưa

PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU

3. Xử Lý Dữ Liệu

Dữ liệu bị thiếu giữa các năm 2014 và 2015.



PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU



3. Xử Lý Dữ Liệu

✓ Vì Weather_description bổ sung loại thông tin với Weather_main

=> Chuyển đổi Weather_main thành các biến được mã hóa một lần và bỏ Weather_description

✓ Thời tiết trong ngày là ngày lễ hoặc cuối tuần

=> Tạo một tính năng mới is_holiday và loại bỏ tính năng cũ ngày lễ không cần phải theo dõi đó là ngày cuối tuần nào sẽ tạo một tính năng mới is_weekend

✓ Cột Date Time rất hữu ích, nhưng không phải ở dạng này.

=> Chuyển đổi nó thành dữ liệu có thể sử dụng là sử dụng sin và cos để chuyển đổi thời gian thành dữ liệu "Time of day" và "Time of year" rõ ràng

PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU



3. Xử Lý Dữ Liệu

temp rain_1h snow_1h clouds_all traffic_volume weather_Clear weather_Clouds weather_Drizzle weather_Fog weather_Haze ... weather_Smoke weather_Snow weather_Squall weather_Thunderstorm is_holiday is_weekend Day sin Day cos Year sin Year cos

date_time

2012-10-02 09:00:00	288.28	0.0	0.0	40	5545	0	1	0	0	0	...	0	0	0	0	0	0	7.071068e-01	-0.707107	-0.999766	0.021642
2012-10-02 10:00:00	289.36	0.0	0.0	75	4516	0	1	0	0	0	...	0	0	0	0	0	0	5.000000e-01	-0.866025	-0.999750	0.022358
2012-10-02 11:00:00	289.58	0.0	0.0	90	4767	0	1	0	0	0	...	0	0	0	0	0	0	2.588190e-01	-0.965926	-0.999734	0.023075
2012-10-02 12:00:00	290.13	0.0	0.0	90	5026	0	1	0	0	0	...	0	0	0	0	0	0	1.143105e-11	-1.000000	-0.999717	0.023791
2012-10-02 13:00:00	291.14	0.0	0.0	75	4918	0	1	0	0	0	...	0	0	0	0	0	0	-2.588190e-01	-0.965926	-0.999700	0.024508
...
2018-09-30 19:00:00	283.45	0.0	0.0	75	3543	0	1	0	0	0	...	0	0	0	0	0	1	-9.659258e-01	0.258819	-0.999910	-0.013421
2018-09-30 20:00:00	282.76	0.0	0.0	90	2781	0	1	0	0	0	...	0	0	0	0	0	1	-8.660254e-01	0.500000	-0.999919	-0.012705
2018-09-30 21:00:00	282.73	0.0	0.0	90	2159	0	0	0	0	0	...	0	0	0	1	0	1	-7.071068e-01	0.707107	-0.999928	-0.011988
2018-09-30 22:00:00	282.09	0.0	0.0	90	1450	0	1	0	0	0	...	0	0	0	0	0	1	-5.000000e-01	0.866025	-0.999936	-0.011271
2018-09-30 23:00:00	282.12	0.0	0.0	90	954	0	1	0	0	0	...	0	0	0	0	0	1	-2.588190e-01	0.965926	-0.999944	-0.010554

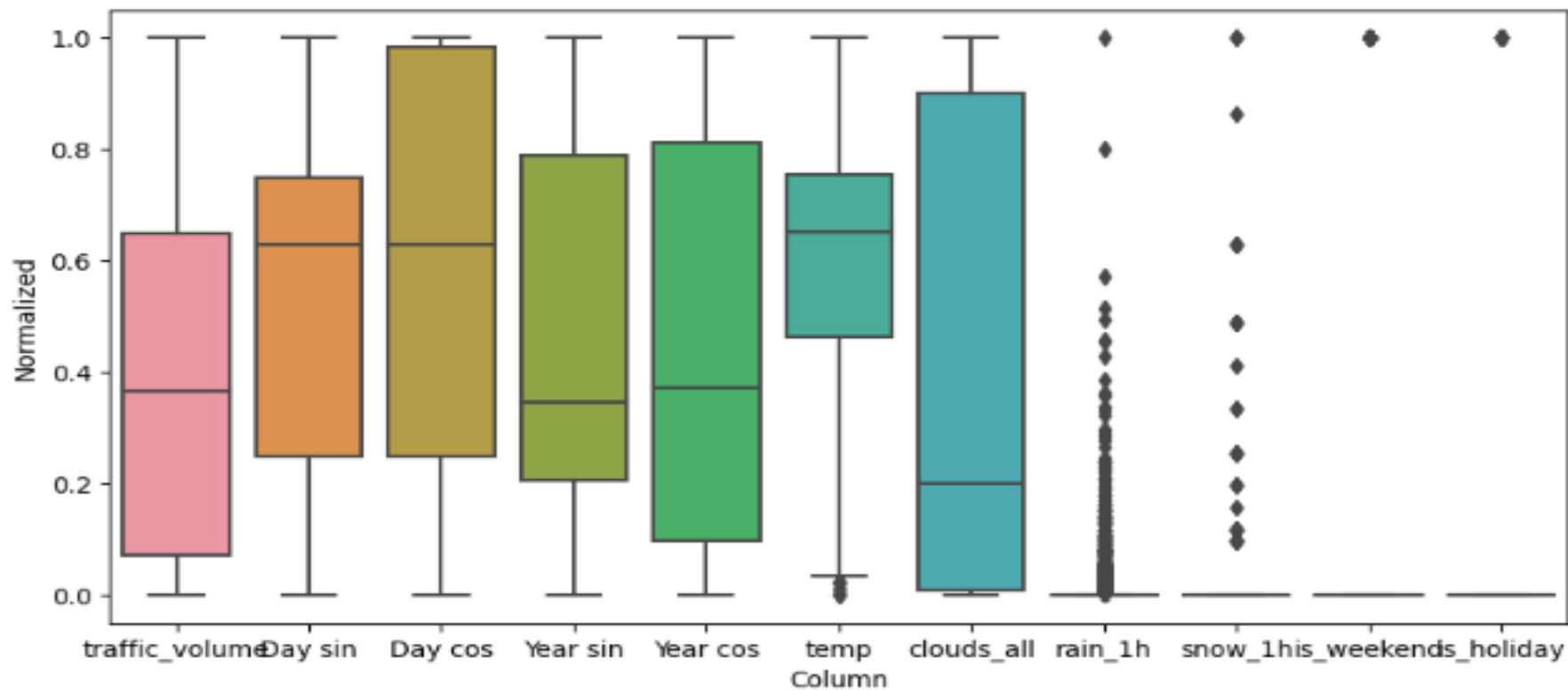
PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU

3. Xử Lý Dữ Liệu

	count	mean	std	min	25%	50%	75%	max
traffic_volume	48204.0	3259.818355	1986.860670	0.00	1193.000000	3.380000e+03	4933.000000	7280.00
Day sin	48204.0	0.014283	0.708420	-1.00	-0.707107	1.525486e-12	0.707107	1.00
Day cos	48204.0	0.006631	0.705630	-1.00	-0.707107	6.790957e-13	0.707107	1.00
Year sin	48204.0	0.012317	0.697627	-1.00	-0.665969	-6.909727e-03	0.706330	1.00
Year cos	48204.0	-0.031720	0.715667	-1.00	-0.744263	-9.217264e-02	0.706604	1.00
temp	48204.0	281.259375	12.712720	243.39	272.160000	2.824500e+02	291.806000	310.07
clouds_all	48204.0	49.362231	39.015750	0.00	1.000000	6.400000e+01	90.000000	100.00
rain_1h	48204.0	0.130315	1.003368	0.00	0.000000	0.000000e+00	0.000000	55.63
snow_1h	48204.0	0.000222	0.008168	0.00	0.000000	0.000000e+00	0.000000	0.51
is_weekend	48204.0	0.284271	0.451071	0.00	0.000000	0.000000e+00	1.000000	1.00
is_holiday	48204.0	0.001265	0.035551	0.00	0.000000	0.000000e+00	0.000000	1.00
weather_Clear	48204.0	0.277799	0.447918	0.00	0.000000	0.000000e+00	1.000000	1.00
weather_Clouds	48204.0	0.314580	0.464353	0.00	0.000000	0.000000e+00	1.000000	1.00
weather_Drizzle	48204.0	0.037777	0.190658	0.00	0.000000	0.000000e+00	0.000000	1.00
weather_Fog	48204.0	0.018920	0.136243	0.00	0.000000	0.000000e+00	0.000000	1.00
weather_Haze	48204.0	0.028213	0.165584	0.00	0.000000	0.000000e+00	0.000000	1.00
weather_Mist	48204.0	0.123434	0.328938	0.00	0.000000	0.000000e+00	0.000000	1.00
weather_Rain	48204.0	0.117667	0.322216	0.00	0.000000	0.000000e+00	0.000000	1.00
weather_Smoke	48204.0	0.000415	0.020365	0.00	0.000000	0.000000e+00	0.000000	1.00
weather_Snow	48204.0	0.059663	0.236864	0.00	0.000000	0.000000e+00	0.000000	1.00
weather_Squall	48204.0	0.000083	0.009109	0.00	0.000000	0.000000e+00	0.000000	1.00
weather_Thunderstorm	48204.0	0.021451	0.144882	0.00	0.000000	0.000000e+00	0.000000	1.00
dayofweek	48204.0	2.984980	2.005941	0.00	1.000000	3.000000e+00	5.000000	6.00
day	48204.0	15.737636	8.722938	1.00	8.000000	1.600000e+01	23.000000	31.00
month	48204.0	6.506037	3.400221	1.00	4.000000	7.000000e+00	9.000000	12.00
year	48204.0	2015.512426	1.893211	2012.00	2014.000000	2.016000e+03	2017.000000	2018.00
day_hour	48204.0	11.398162	6.940238	0.00	5.000000	1.100000e+01	17.000000	23.00

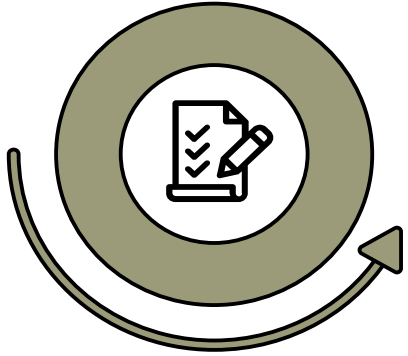
PHÂN TÍCH VÀ XỬ LÝ DỮ LIỆU

3. Xử Lý Dữ Liệu



=> Dữ liệu cho thấy mưa và tuyết có rất nhiều điểm bất thường nhưng đây là do điều kiện thời tiết ở khu vực đó.

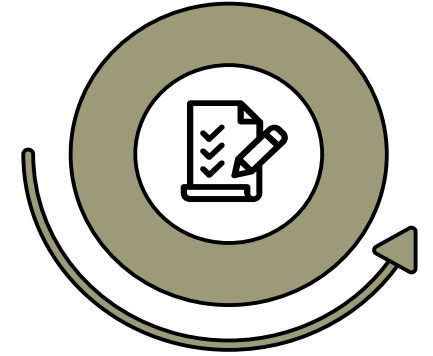
KẾT QUẢ XÂY DỰNG MÔ HÌNH



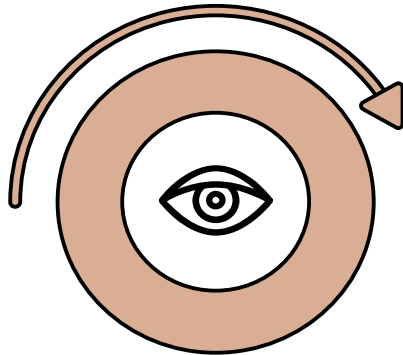
1. Mô Hình Dense



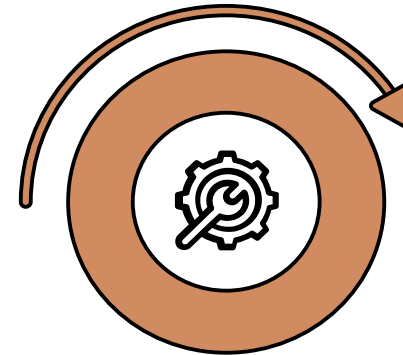
3. Mô Hình LSTM



5. So Sánh Hiệu Suất Mô Hình



2. Mô Hình Tích Chập
Conv(CNN)

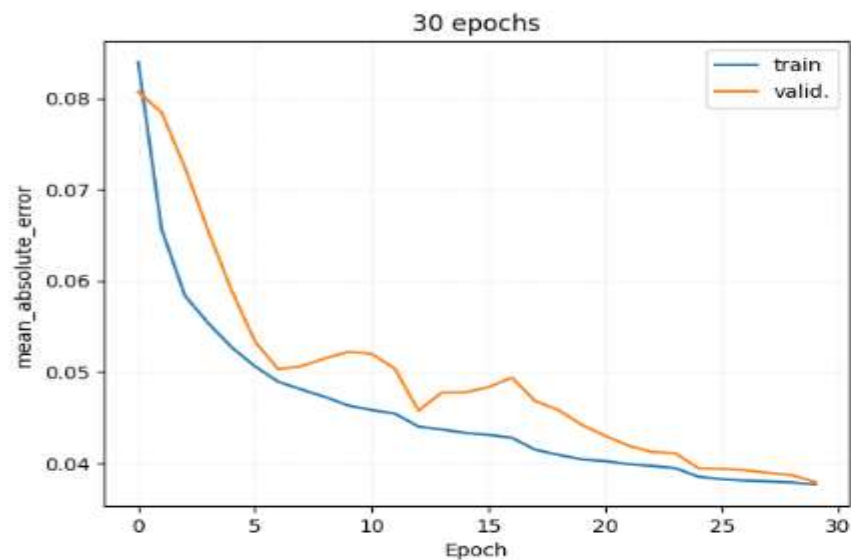


4. Mô Hình My Models



Mô Hình Dense

Training vs. Validation:

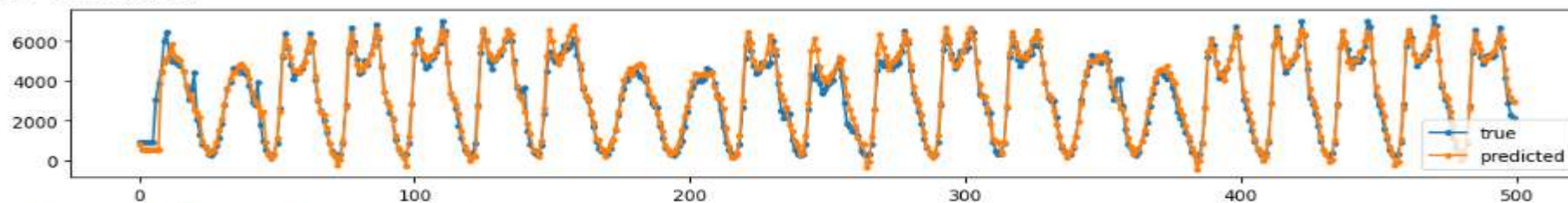


Validation Scores:

```
261/261 - 1s - loss: 0.0031 - mean_absolute_error: 0.0379 - 1s/epoch - 4ms/step  
261/261 [=====] - 1s 4ms/step
```

Predictions Evaluation:

```
Predictions: 8352  
MAE: 276.05 (0.0379)
```

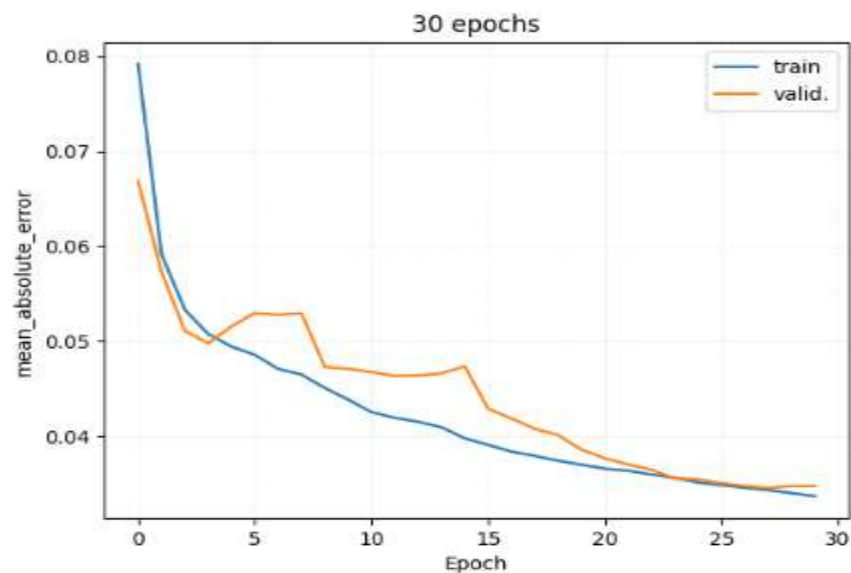


```
CPU times: user 5min 41s, sys: 48.5 s, total: 6min 29s  
Wall time: 5min 40s
```



Mô Hình Tích Chập Conv(CNN)

Training vs. Validation:

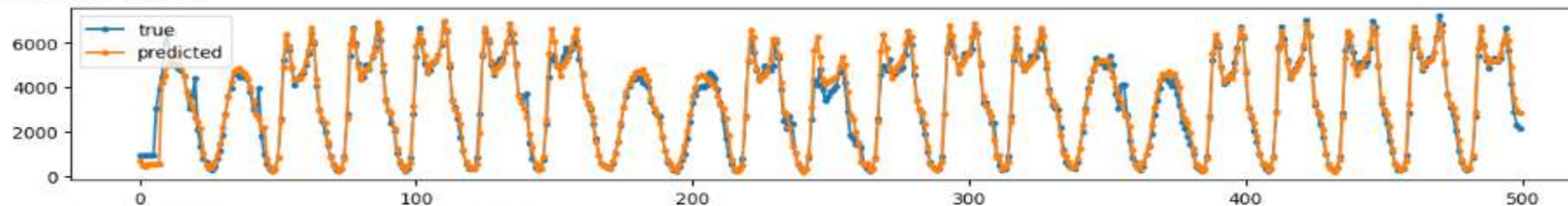


Validation Scores:

261/261 - 1s - loss: 0.0029 - mean_absolute_error: 0.0348 - 1s/epoch - 4ms/step
261/261 [=====] - 2s 6ms/step

Predictions Evaluation:

Predictions: 8352
MAE: 253.03 (0.0348)

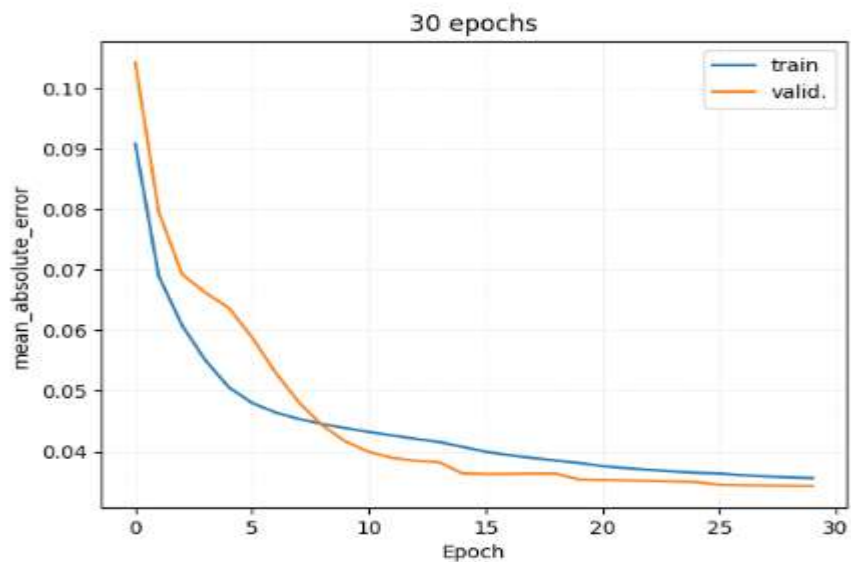


CPU times: user 5min 44s, sys: 47.1 s, total: 6min 31s



Mô Hình LSTM

Training vs. Validation:

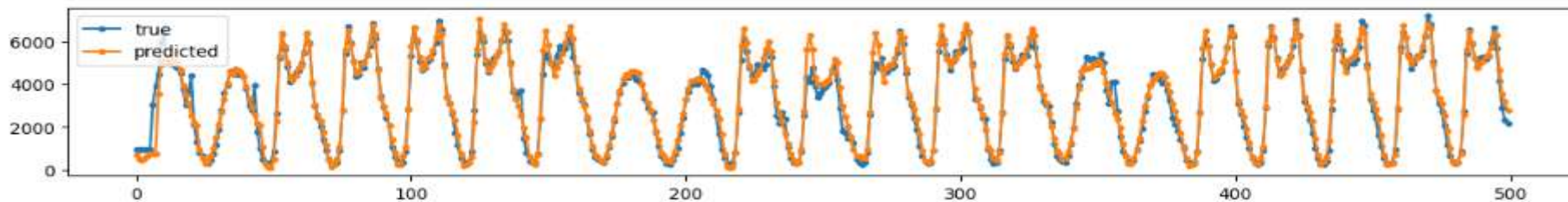


Validation Scores:

261/261 - 1s - loss: 0.0027 - mean_absolute_error: 0.0343 - 956ms/epoch - 4ms/step
261/261 [=====] - 1s 3ms/step

Predictions Evaluation:

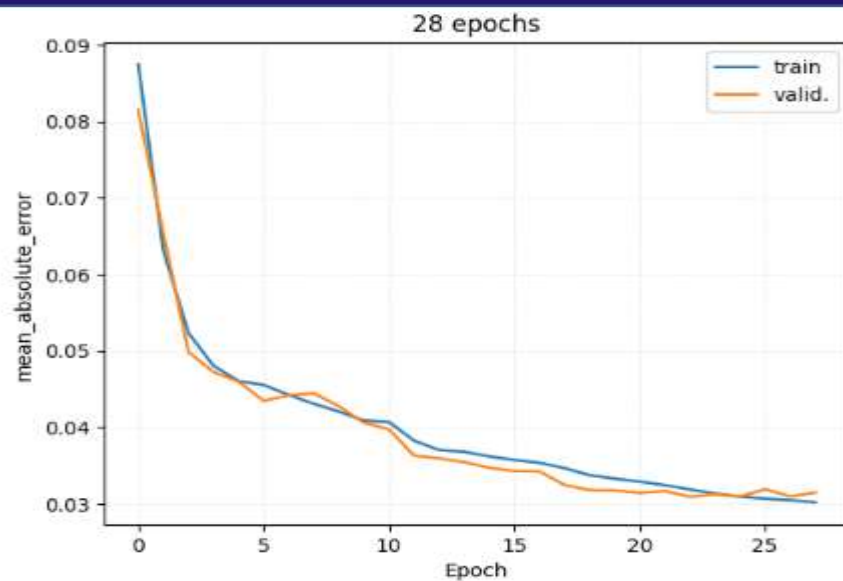
Predictions: 8352
MAE: 249.86 (0.0343)



CPU times: user 6min 5s, sys: 45.5 s, total: 6min 51s
Wall time: 5min 13s



Mô Hình My Models - 1

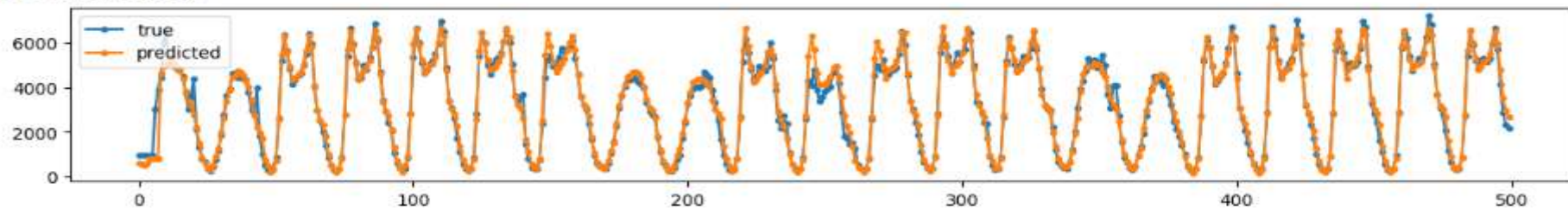


Validation Scores:

261/261 - 1s - loss: 0.0023 - mean_absolute_error: 0.0309 - 1s/epoch - 5ms/step
261/261 [=====] - 2s 4ms/step

Predictions Evaluation:

Predictions: 8352
MAE: 224.84 (0.0309)

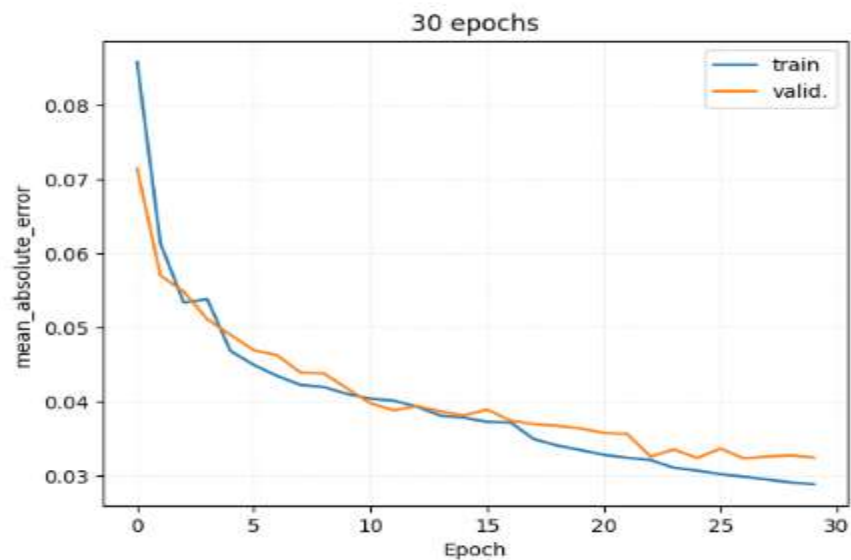


CPU times: user 7min 21s, sys: 43.1 s, total: 8min 4s
Wall time: 7min 3s



Mô Hình My Models - 2

Training vs. Validation:

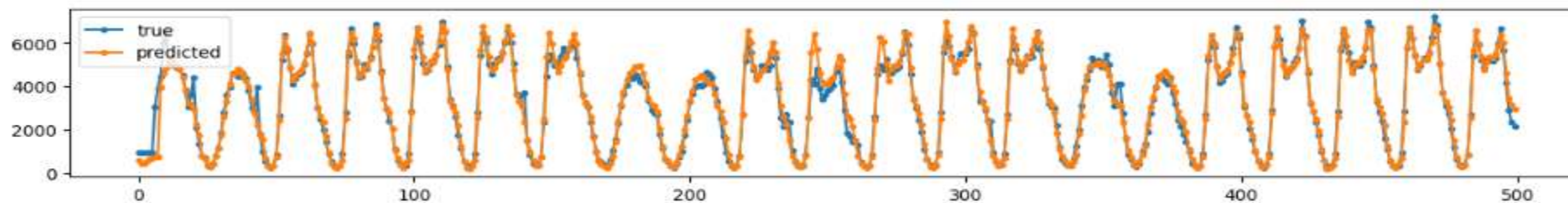


Validation Scores:

261/261 - 2s - loss: 0.0025 - mean_absolute_error: 0.0324 - 2s/epoch - 9ms/step
261/261 [=====] - 3s 7ms/step

Predictions Evaluation:

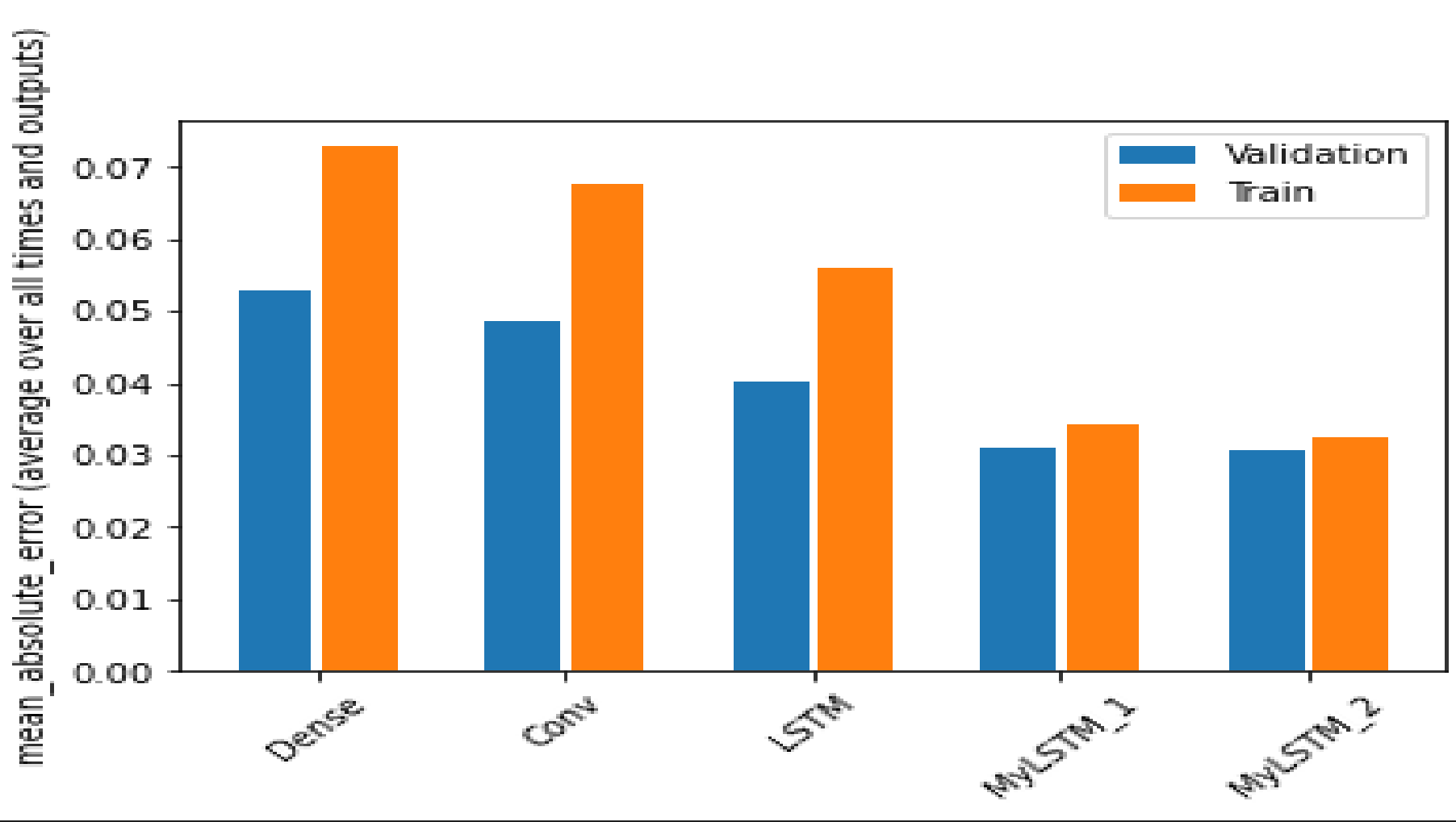
Predictions: 8352
MAE: 235.53 (0.0324)



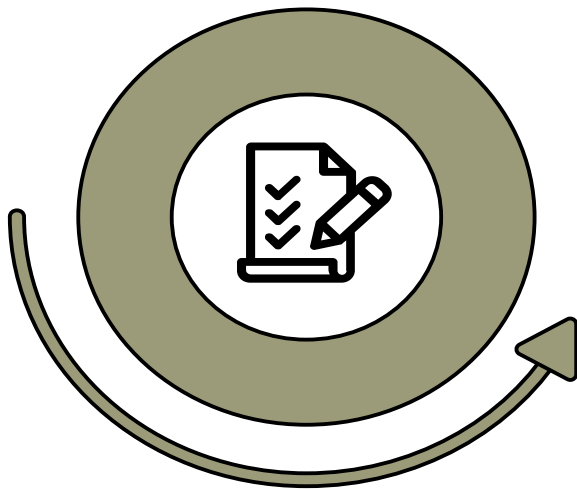
CPU times: user 11min 18s, sys: 57.2 s, total: 12min 15s
Wall time: 10min 35s



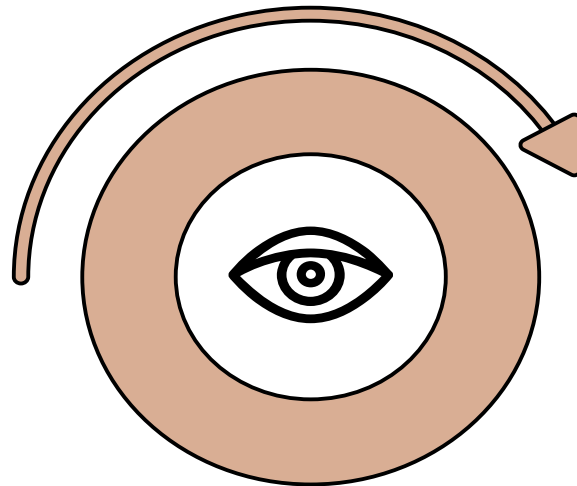
So Sánh Hiệu Suất Mô Hình



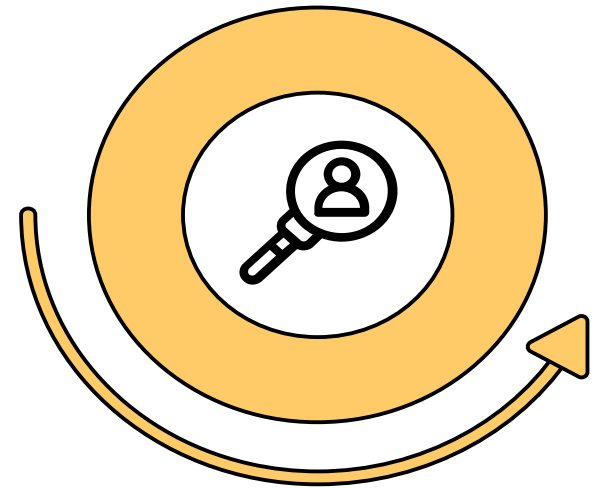
KẾT LUẬN



1. Kết Quả Đạt Được




2. Hạn Chế



3. Hướng Phát Triển

KẾT LUẬN



1. Kết Quả Đạt Được

- ✓ Hyperparameter đã được kiểm tra một cách có hệ thống bằng các giá trị khác nhau là:
 - Số lượng các đơn vị và lớp LSTM và các đơn vị dày đặc ở các lớp cuối cùng (Mô hình 1)
 - LSTM hai chiều (Mô hình 2)
- ✓ Hơn nữa, em đã giữ một biến thể của ba mô hình tham chiếu được cung cấp trong hướng dẫn Tensorflow (Dense, Conv và LSTM). Em đã xây dựng hai mô hình khác (mylstm_1, mylstm_2) hoạt động tốt hơn so với các mô hình tham chiếu.
- ✓ Mô hình tốt nhất của em sử dụng các LSTM hai chiều với hai lớp chuyển tiếp và lùi tùy chỉnh, và hai lớp dày đặc với 512 đơn vị mỗi lớp và một lớp đầu ra dày đặc với một đơn vị. Mô hình đã thực hiện phương sai thấp nhất và được duy trì cho nhiều kỷ nguyên hơn trong quá trình đào tạo
- ✓ Các dự đoán trong (Hình) cho thấy mô hình đã nắm bắt được tất cả các mẫu quan trọng với các lần bỏ lỡ nhỏ của một số dị thường.

KẾT LUẬN



2. Hạn Chế

- ✓ Dự đoán lưu lượng giao thông trên bộ dữ liệu trực tuyến có thể Không hoàn toàn chính xác một cách hoàn hảo
- ✓ Khi phân tích và xử lý dữ liệu trực tuyến, việc thu thập và xử lý dữ liệu có thể gặp độ trễ.
- ✓ Sự biến đổi không đồng đều, Lưu lượng giao thông có thể thay đổi theo thời gian và địa điểm.

KẾT LUẬN

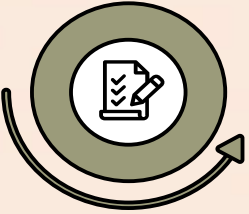


3. Hướng Phát Triển

- ✓ Mô hình có thể áp dụng dựa trên tập dữ liệu và các tuyến đường giao thông ở Việt Nam. Tích hợp hệ thống quản lý giao thông thông minh, đề tài có thể được mở rộng để tích hợp với hệ thống quản lý giao thông thông minh (ITS) để tạo ra giải pháp toàn diện.
- ✓ Phân tích và dự đoán thời gian hành trình, ngoài việc dự đoán lưu lượng giao thông, cũng có thể mở rộng để phân tích và dự đoán thời gian hành trình của các phương tiện. Điều này có thể giúp người dùng và hệ thống đưa ra các quyết định thông minh về lộ trình và thời gian đi lại.



Type here to search



Type here to search



13:10 PM
09/05/2022