

Chương 3.

MÔ HÌNH DỰ ĐOÁN CHUỖI THỜI GIAN ĐƠN BIẾN

PhD. Nguyễn Thị Khánh Tiên

email: tienntk@ut.edu.vn



Nội dung chính trong chương 3:

1. Chuỗi thời gian đơn biến và các đặc trưng quan trọng
2. Nhóm mô hình dự đoán chuỗi thời gian đơn biến
 - Mô hình thống kê cổ điển: mô hình tự hồi quy AR, mô hình trung bình trượt MA, mô hình ARIMA, SARIMA
 - Mô hình Machine Learning: Linear Regression, Random Forest, XGBoost
 - Mô hình Deep Learning

Mục tiêu thực hành chương 3:

1. Hiểu rõ bản chất và đặc trưng của chuỗi thời gian đơn biến trong bài toán dự báo.
2. Nắm vững nguyên lý hoạt động của các nhóm mô hình dự đoán chuỗi thời gian
3. Phân tích được điều kiện áp dụng, ưu điểm và hạn chế của từng nhóm mô hình.
4. Biết cách lựa chọn mô hình phù hợp dựa trên đặc tính dữ liệu và mục tiêu bài toán.
5. Chuẩn bị nền tảng lý thuyết cho các chương thực hành dự báo chuỗi thời gian nâng cao.

Khái niệm chuỗi thời gian đơn biến (Univariate Time Series)

Chuỗi thời gian đơn biến (Univariate Time Series) là một dãy các giá trị quan sát của **một biến duy nhất**, được ghi nhận theo **thứ tự thời gian** và thường cách đều nhau.

Ký hiệu:

$$\{y_t\} = y_1, y_2, \dots, y_T$$

Trong đó:

- y_t : giá trị của biến tại thời điểm t ,
- T : tổng số quan sát.

Ví dụ:

- Doanh thu theo tháng
- Nhiệt độ trung bình theo ngày
- Số lượng hành khách theo năm
- Giá cổ phiếu theo phiên giao dịch

Đặc điểm quan trọng của chuỗi thời gian là **tính phụ thuộc theo thời gian**:

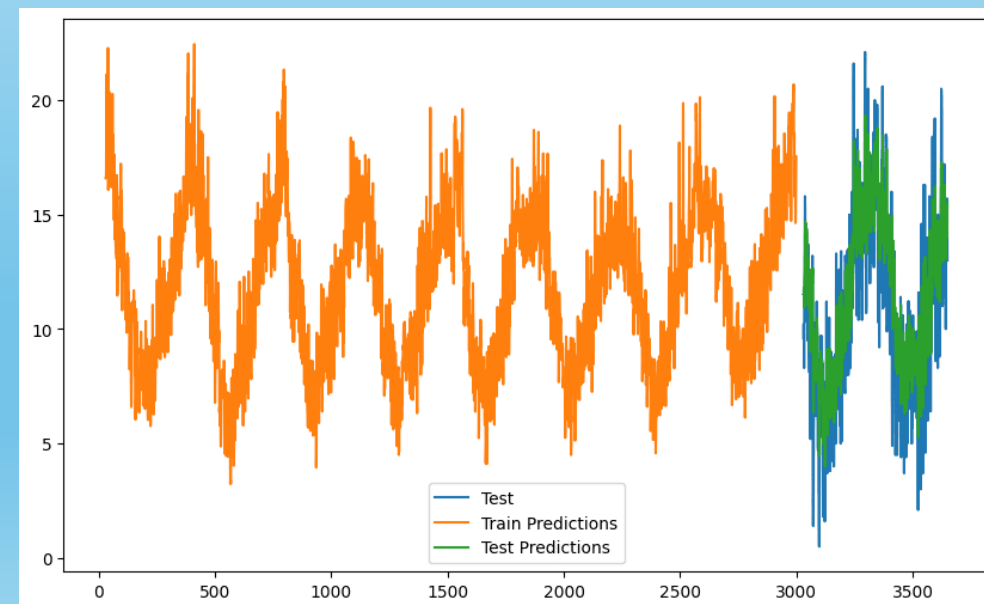
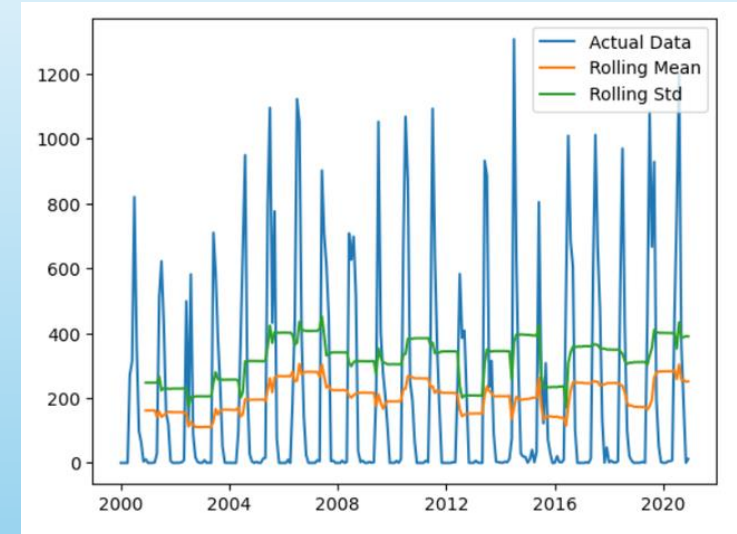
- giá trị tại thời điểm hiện tại thường phụ thuộc vào các giá trị trong quá khứ.

Chuỗi thời gian đơn biến:

- Có cấu trúc phụ thuộc theo thời gian.
- Gồm nhiều thành phần (trend, seasonality, noise).
- Có các đặc trưng quan trọng như tính dừng, tự tương quan, độ trễ.

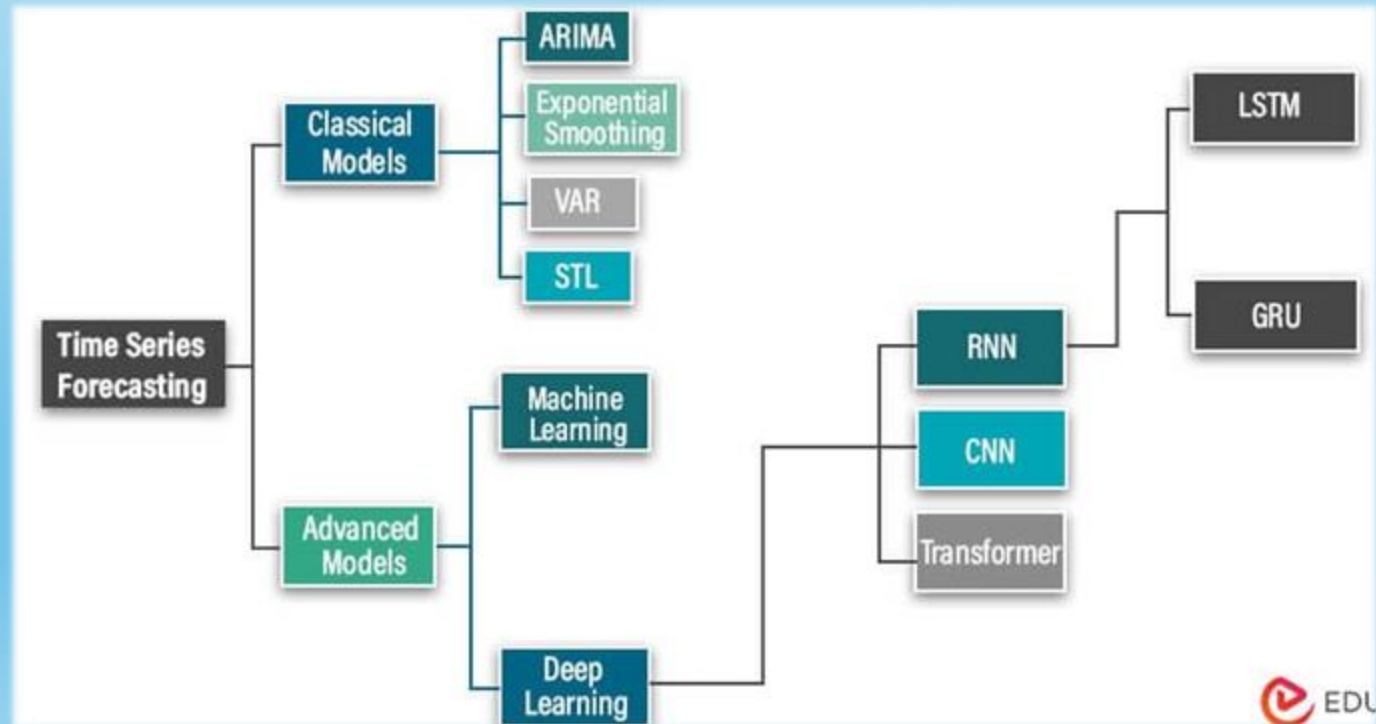
Tính dừng (Stationarity). Một chuỗi được gọi là **dừng** nếu:

- Kỳ vọng (mean) không đổi theo thời gian.
- Phương sai không đổi.
- Cấu trúc tự tương quan không thay đổi theo thời gian.
- Tính dừng là giả định quan trọng của nhiều mô hình thống kê



Nhóm mô hình dự đoán chuỗi thời gian đơn biến

| Nhóm mô hình | Ưu điểm | Nhược điểm | Nên dùng khi |
|------------------|----------------------------------|-----------------|---------------------------|
| Thống kê | Dễ diễn giải, lý thuyết chặt chẽ | Giả định mạnh | Chuỗi ngắn, ổn định |
| Machine learning | Linh hoạt, hiệu quả cao | Cần feature tốt | Chuỗi vừa, phi tuyến tính |
| Deep Learning | Học đặc trưng mạnh | Tốn dữ liệu | Chuỗi lớn, phức tạp |



Mô hình thống kê cổ điển

Các mô hình:

- Mô hình tự hồi quy (AR – Autoregressive)
- Mô hình trung bình trượt (MA – Moving Average)
- Mô hình ARIMA (AR + I + MA)
- Mô hình SARIMA (mở rộng ARIMA cho dữ liệu có mùa vụ)

Các vấn đề:

- Ý nghĩa các tham số (p, d, q) và (P, D, Q, s).
- Vai trò của ACF và PACF trong lựa chọn mô hình.
- Điều kiện áp dụng: chuỗi dừng, tuyến tính, giả định phân phối.
- Ưu điểm và hạn chế của các mô hình thống kê cổ điển.

Mục tiêu:

- Hiểu nguyên lý hoạt động của các mô hình truyền thống.
- Biết khi nào nên sử dụng ARIMA/SARIMA.
- Làm cơ sở so sánh với các mô hình Machine Learning.
- định dạng).

Giới thiệu thư viện statsmodels

statsmodels là thư viện Python chuyên về thống kê, kinh tế lượng và phân tích chuỗi thời gian.

Thư viện này cung cấp các mô hình có nền tảng toán học chặt chẽ, cho phép:

- Ước lượng tham số theo lý thuyết thống kê
- Kiểm định giả thuyết
- Phân tích phần dư
- Diễn giải kết quả mô hình

Trong phân tích chuỗi thời gian, statsmodels được xem là **thư viện chuẩn (standard library)** cho các mô hình thống kê cổ điển

Ref:

<https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>

https://www.statsmodels.org/stable/generated/statsmodels.tsa.ar_model.AutoReg.html

1. Basic Time Series Operations with Pandas

- Convert to DateTime: `pd.to_datetime(df['date_column'])`
- Set DateTime as Index: `df.set_index('datetime_column', inplace=True)`
- Resample Time Series Data: `df.resample('D').mean()`
- Time-Based Indexing: `df['2020-01-01':'2020-01-31']`
- Rolling Window Statistics: `df.rolling(window=5).mean()`
- Expanding Window Statistics: `df.expanding(min_periods=1).mean()`
- Shifting and Lagging: `df.shift(1)`
- Differencing (for Stationarity): `df.diff(periods=1)`

2. Visualization of Time Series Data

- Line Plot: `df.plot()`
- Rolling Mean Plot: `df.rolling(window=12).mean().plot()`
- Histogram and Density Plot: `df.hist(), df.plot(kind='kde')`
- Box Plot of Yearly/Monthly Data: `df.boxplot(groupby=df.index.year)`

3. Seasonal Decomposition and Analysis

- Seasonal Decomposition with statsmodels: `from statsmodels.tsa.seasonal import seasonal_decompose; seasonal_decompose(df, model='additive')`
- Autocorrelation and Partial Autocorrelation Plots: `from statsmodels.graphics.tsaplots import plot_acf, plot_pacf; plot_acf(df)`

4. Time Series Forecasting Models

- ARIMA Model with statsmodels: `from statsmodels.tsa.arima.model import ARIMA; ARIMA(df, order=(5,1,0)).fit()`
- Seasonal ARIMA (SARIMA): `from statsmodels.tsa.statespace.sarimax import SARIMAX; SARIMAX(df, order=(1, 1, 1), seasonal_order=(1, 1, 1, 12)).fit()`

Giới thiệu thư viện statsmodels

| Nhóm | Công cụ / Mô hình | Chức năng chính | Mục đích sử dụng |
|--------------------|-------------------|---|------------------------------|
| Khám phá dữ liệu | ACF | Đo tương quan giữa (y_t) và (y_{t-k}) | Phát hiện cấu trúc phụ thuộc |
| | PACF | Tương quan riêng phần theo lag | Chọn bậc p cho AR |
| | Decomposition | Phân rã trend, seasonality, residual | Hiểu cấu trúc chuỗi |
| Kiểm tra tính dừng | ADF | Kiểm định nghiệm đơn vị | Xác định chuỗi có dừng |
| | KPSS | Kiểm định tính dừng | Kiểm tra bổ sung cho ADF |
| Mô hình | AR | Tự hồi quy theo lag | Chuỗi dừng, không mùa vụ |
| | MA | Trung bình trượt nhiều | Chuỗi có cú sốc ngắn hạn |
| | ARIMA | AR + sai phân + MA | Chuỗi không dừng |
| | SARIMA / SARIMAX | ARIMA mùa vụ (\pm biến ngoại sinh) | Chuỗi có mùa vụ |
| | | | |
| Đánh giá & Dự báo | AIC | Độ phù hợp mô hình | So sánh mô hình |
| | BIC | Phạt độ phức tạp | Chọn mô hình gọn |
| | Residual Analysis | Phân tích phần dư | Kiểm tra giả định |
| | Forecast | Dự báo tương lai | Ứng dụng thực tế |

Reference

- <https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>