# Dữ liệu chuỗi thời gian

Dữ liệu chuỗi thời gian là tập hợp các quan sát thu được thông qua các phép đo lặp lại theo thời gian. Vẽ các điểm biểu diễn trên đồ và một trong các trục sẽ luôn là thời gian.

Số liệu chuỗi thời gian đề cập đến một phần dữ liệu được dõi theo thời gian. Ví dụ: Số liệu có thể đề cập đến số lượng hàng tồn kho đã bán dược trong cửa hàng từ ngày này sang ngày khác.

Dữ liệu chuối thời gian có ở khắp mọi nơi vì thời gian là thành phần của mọi thứ có thể quan sát được. Khi thế giới ngày càng được trang bị nhiều thiaats bị, các cảm biến và hệ thống liên tục phát ra dòng dữ liệu chuỗi thời gian không ngừng. Dữ liệu như vậy có nhiều ứng dụng trên nhiều ngành công ngiệp khác nhau. Hãy đặt điều này vào bối cảnh thông qua một số ví dụ.

Ví dụ về phân tích chuỗi thời gian:

- Hoạt động điện trong não.
- Đo lượng mưa.
- Giá cổ phiếu.
- Nhịp tim mỗi phút…

# Thành phần chuỗi thời gian

Một cách trừu tượng hữu ích để lựa các phương pháp dự báo là chia chuỗi thời gian thành các thành phần có hệ thống và không có hệ thống.

- Có hệ thống: Các hành phần của chuỗi thời gian có tính nhất quán hoặc lặp lại và có thể được mô tả và mô hình hoá.
- Không có hệ thống: Các thành phần của chuỗi thời gian không thể được mô hình hoá trực tiếp.

Một chuỗi thời nhất định được cho là bao gồm ba thành phần mang tính hệ thống bao gồm mức độ, xu hướng, tính thời vụ và một thành phần không mang tính hệ thống đó là độ nhiễu.

Các thành phần này được xác định như sau:

- Cấp độ: Giá trị trung bình trong chuỗi.
- Trend: Giá trị tăng hoặc giảm trong chuỗi.
- Tính thời vụ: Chu kỳ ngắn hạn lặp đi lặp lại trong chuỗi.
- Độ nhiễu: Sự biến đổi ngẫu nhiên trong chuỗi.

# Phân rã chuỗi thời gian

Phân ra chuỗi thời gian liên quan đến việc coi chuỗi là sự kết hợp của các thành phần cấp độ, xu hướng, tính thời vụ và độ nhiễu.

Sự phân rã cung cấp một mô hình trừu tượng hữu ích để suy nghĩ về chuỗi thời gian nói chung và để. hiểu rõ hơn các vấn đề trong quá trình phân tích dự báo chuỗi thời gian.

# Các khái niệm cơ bản trong mô hình chuỗi thời gian

Một hình chuỗi thời gian được sử dụng vì nhiều lý do, dự đoán kết quả trong tương lai, hiểu kết quả trong quá khứ, đưa ra đề xuất và các chính sách,... Các mục tiêu chung này của việc lập mô hình dữ liệu chéo hoặc dữ liệu bảng. Tuy nhiên, các kỹ thuật được sử dụng trong mô hình chuỗi thời gian phải tính đến mối tương quan chuỗi thời gian.

Trong mô hình chuỗi thời gian cơ bản sau sẽ giúp ta hình dung rõ ràng hơn về một mô hình chuỗi thời gian hoàn chỉnh:

- Mô hình miền thời gian và miền tần số.
- Mô hình chuỗi thời gian đơn biến và đa biến.
- Mô hình chuỗi thời gian tuyến tính và phi tuyến tính.

# Pandas và Python trong dữ liệu chuỗi thời gian

Pandas chứa các khả năng và tính năng mở rộng để làm việc với dữ liêu chuỗi thời gian cho tất cả các miền. Bằng cách sử dụng NumPy datetime64 và timedelta64 dtypes, pandas đã hợp nhất một số lượng lớn các tính năng từ các thư viện Python khác cũng như scikits.timeseries tạo ra một lượng lớn chức năng để thao tác dữ liệu chuỗi thời gian.

Mặc dù chuỗi thời gian cũng có sẵn scikit-learn nhưng Pandas có một số tính năng được biên soạn nhiều hơn. Trong model Pandas này, có thể bao gồm ngày và giờ cho mỗi bản ghi và có thể tìm nạp các bản ghi của khung dữ liệu. Chúng ta có thể tìm ra dữ liệu trong một phạm vi ngày và giờ nhất định bằng cách sử dụng model Pandas có tên Time series.

Mục tiêu của phân tích chuỗi thời gian với Pandas:

- Tạo chuỗi ngày.
- Làm việc với dấu thời gian dữ liệu.
- Chuyển đổi dữ liệu chuỗi thời gian thành dấu thời gian.

- Cắt dữ liệu chuỗi thời gian bằng dấu thời gian.
- Lấy mẫu lại chuỗi thời gian của bạn cho các tổng hợp/thống kê tóm tắt trong khoảng thời gian khác nhau.
- Làm việc với dữ liệu bị thiếu

Sau đây là có ví dụ về pandas:

Code 1:

```
import pandas as pd
from datetime import datetime
import numpy as np

range_date = pd.date_range(start ='1//1/2019', end = '1/08/2019',
                           freq='Min')
range_date
```

✓ 0.0s Python

Output:

```
DatetimeIndex(['2019-01-01 00:00:00', '2019-01-01 00:01:00',
               '2019-01-01 00:02:00', '2019-01-01 00:03:00',
               '2019-01-01 00:04:00', '2019-01-01 00:05:00',
               '2019-01-01 00:06:00', '2019-01-01 00:07:00',
               '2019-01-01 00:08:00', '2019-01-01 00:09:00',
               ...
               '2019-01-07 23:51:00', '2019-01-07 23:52:00',
               '2019-01-07 23:53:00', '2019-01-07 23:54:00',
               '2019-01-07 23:55:00', '2019-01-07 23:56:00',
               '2019-01-07 23:57:00', '2019-01-07 23:58:00',
               '2019-01-07 23:59:00', '2019-01-08 00:00:00'],
              dtype='datetime64[ns]', length=10081, freq='T')
```

Giải thích:

Với code trên, đã tạo ra dấu thời gian dựa trên số phút cho các phạm vi từ 1/1/2019-1/08/2019. Chúng ta có thể thay đổi tần số theo giờ, phút hoặc giây. Chức năng này sẽ giúp bạn theo dõi bản ghi dữ liệu được lưu trữ mỗi phút. Như chũng ta có thể thấy ở đầu ra, độ dài của tem datetime là 10081. Hãy nhớ rằng cấu trúc sử dụng kiểu datetime64[ns].

Code2:

```
import pandas as pd
from datetime import datetime
import numpy as np

range_date = pd.date_range(start ='1//1/2019', end = '1/08/2019',
                           freq='Min')
df = pd.DataFrame(range_date, columns = ['date'])
df['data'] = np.random.randint(0, 100, size = len(range_date))

print(df.head(10))
```

✓ 0.0s Python

Output:

```
                 date  data
0 2019-01-01 00:00:00    91
1 2019-01-01 00:01:00    21
2 2019-01-01 00:02:00    47
3 2019-01-01 00:03:00    66
4 2019-01-01 00:04:00    78
5 2019-01-01 00:05:00    87
6 2019-01-01 00:06:00    89
7 2019-01-01 00:07:00    60
8 2019-01-01 00:08:00    11
9 2019-01-01 00:09:00    45
```

Giải thích:

Trước tiên, chúng ta sẽ tạo ra chuôi thời gian sau đó chuyển đổi dữ liệu này thành khung dữ liệu và sử dụng hàm ngẫu nhiên để tạo dữ liệu ngẫu nhiên và ánh xạ trên khung dữ liệu. Sau đó để kiểm tra kết quả chúng ta sử dụng hàm print.

Để thực hiện các thao tác chuỗi thời gian, chũng ta cần có chỉ mục ngày giờ đề khung dữ liệu được lập chỉ mục trên dấu thời gian. Ở đây, chúng ta đang thêm một cột mới nữa khung dữ liệu và cấu trúc.