# Time Series Analysis - Key Concepts

## ****1. LSTM (Long Short-Term Memory)****

LSTM là một loại mạng nơ-ron hồ tiếp (**Recurrent Neural Network - RNN**) được thiết kế để xử lý **dữ liệu chuỗi thời gian dài**. Điểm khác biệt chính của LSTM so với RNN truyền thống là nó sử dụng các **"cổng kiểm soát"** để điều chỉnh thông tin được lưu giữ hoặc loại bỏ qua từng bước thời gian. Điều này giúp LSTM **"giảm thiểu vấn đề vanishing gradient"**, làm cho nó đặc biệt hiệu quả khi xử lý các chuỗi dài và phức tạp.

## ****2. GRU (Gated Recurrent Unit)****

GRU là một phiên bản **"tinh gọn"** của LSTM, vẫn giữ được khả năng **"ghi nhớ thông tin dài hạn"** nhưng sử dụng ít tham số hơn. Thay vì ba cổng như LSTM, GRU chỉ có **"reset gate"** và **"update gate"**, giúp quá trình huấn luyện nhanh hơn và ít tốn tài nguyên hơn. Nếu dữ liệu có tính tuần hoàn và xu hướng rõ ràng, GRU có thể là một lựa chọn hiệu quả hơn so với LSTM.

## ****3. Dense Layer****

Dense Layer là một lớp trong mạng nơ-ron mà **"mỗi neuron trong lớp này kết nối đầy đủ với tất cả neuron của lớp trước đó"**. Điều này giúp mô hình có khả năng học tốt hơn các biểu diễn phức tạp của dữ liệu. Trong các mạng deep learning, **"Dense Layer"** đóng vai trò quan trọng trong việc **"tổng hợp thông tin từ các tầng trước đó"**, đặc biệt là sau các lớp CNN hoặc RNN.

## ****4. Dropout****

Dropout là một kỹ thuật **"regularization"** phổ biến trong deep learning, giúp **"giảm overfitting"** bằng cách **"tắt ngẫu nhiên một số neuron trong quá trình huấn luyện"**. Điều này làm cho mô hình **"bớt phụ thuộc vào một số đặc trưng cụ thể"**, từ đó tổng quát hóa tốt hơn trên dữ liệu thực tế. Khi sử dụng Dropout, mỗi lần huấn luyện, một tỷ lệ % các neuron sẽ bị **"bỏ qua"**, giúp mô hình học cách dự đoán một cách linh hoạt hơn.

## ****5. Resampling****

Resampling là kỹ thuật **"thay đổi tần suất dữ liệu chuỗi thời gian"**, chẳng hạn như **"chuyển đổi dữ liệu từ hàng ngày thành hàng tuần"** hoặc ngược lại. Kỹ thuật này giúp làm **"mịn dữ liệu"**, xử lý các giá trị bị thiếu và tạo ra một tập dữ liệu phù hợp hơn với mô hình dự báo.

## ****6. Upsampling****

Upsampling là một phương pháp **"tăng cường tần suất dữ liệu"** bằng cách tạo thêm điểm dữ liệu ở khoảng thời gian ngắn hơn. Nó thường được sử dụng khi dữ liệu gốc quá thưa thớt và cần **"nội suy"** để có thêm thông tin. Ví dụ, khi làm việc với dữ liệu tài chính theo tháng, có thể cần **"tăng độ phân giải dữ liệu"** để dự báo hàng ngày một cách chính xác hơn.

## ****7. Growth Rates****

Growth Rate là cách đo lường **"tốc độ thay đổi"** của một chuỗi thời gian theo thời gian, thường xuất hiện trong **"phân tích tài chính, kinh tế và khoa học dữ liệu"**. Nó giúp trả lời câu hỏi như **"doanh thu công ty này đang tăng trưởng bao nhiêu phần trăm so với năm ngoái?"**. Công thức tính đơn giản là **"tỷ lệ thay đổi giữa hai thời điểm"**, thường được biểu thị dưới dạng phần trăm (%).

## ****8. Lag****

Lag đề cập đến **"sự trễ giữa các quan sát"** trong chuỗi thời gian. Đây là một khái niệm quan trọng khi phân tích mối quan hệ giữa dữ liệu hiện tại và dữ liệu quá khứ. Ví dụ, nếu doanh số bán hàng của một công ty có **"độ trễ 3 tháng so với ngân sách quảng cáo"**, tức là tác động của quảng cáo hôm nay sẽ chỉ ảnh hưởng đến doanh số sau 3 tháng. **"Lag variables"** thường được sử dụng trong mô hình hồi quy để phát hiện xu hướng hoặc độ trễ trong dữ liệu.

## ****9. Changes (Time Series)****

Changes trong chuỗi thời gian là sự **"chênh lệch giữa các giá trị liên tiếp"**, giúp phát hiện xu hướng và mức độ biến động của dữ liệu. Nếu một chuỗi thời gian có **"sự thay đổi lớn giữa các điểm dữ liệu"**, có thể nó đang chịu tác động bởi một sự kiện đặc biệt, như một chính sách mới hoặc biến động thị trường.

## ****10. Autoregression (AR)****

Autoregression (AR) là một mô hình thống kê sử dụng **"các giá trị trong quá khứ"** để dự đoán giá trị tương lai trong chuỗi thời gian. **"Mô hình AR"** giả định rằng **"mối quan hệ giữa dữ liệu hiện tại và dữ liệu trước đó là tuyến tính"**, tức là giá trị tại thời điểm hiện tại có thể được biểu diễn bằng tổ hợp tuyến tính của các giá trị trước đó. Đây là nền tảng cho nhiều mô hình dự báo chuỗi thời gian như **"ARIMA"**.