ỦY BAN NHÂN DÂN TP HỒ CHÍ MINH

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO ĐỒ ÁN**

ĐỀ TÀI: DỰ ĐOÁN LƯỢNG

TRUY CẬP TRANG WEB BẰNG MÔ HÌNH ARIMA

Giảng viên hướng dẫn: **Ts. Trịnh Tấn Đạt**

Sinh viên thực hiện: **Trần Ngọc Vũ**

MSSV: **3121410581**

**NIÊN KHÓA: 2021 - 2025**

**MỤC LỤC**

Lời nói đầu

Giới thiệu đề tài

1. Tổng quan về đề tài
2. Mô tả bài toán

Chuỗi thời gian

Mô hình ARIMA

1. Lý thuyết
   1. Auto regression
   2. Moving average
   3. Intergated
2. Xây dựng mô hình
   1. Khảo sát tính dừng
   2. ACF và PACF
   3. Lựa chọn tham số cho mô hình ARIMA
3. Khảo sát
   1. Đọc hiểu kết quả mô hình ARIMA

Dự báo

Đánh giá

Tổng kết

**LỜI NÓI ĐẦU**

Trong bối cảnh internet phát triển như ngày nay, hiện tượng các website mọc lên như nấm sau mưa là điều hiển nhiên. Việc quản lí các webstite là việc làm cần thiết không thể thiếu, biết rõ số lượng khách hàng truy cập mà một trong những việc làm đó. Việc nắm được lượng truy cập của khách hàng là việc làm cần thiết để tối ưu tiền bạc và sức lực đặc biệt là các website về giải trí và kinh doanh. Ví dụ trong các dịp lễ hay các mùa giảm giá biết được lượng khách hàng truy cập, từ đó có các chiến lược marketing ít tồn tiền bạc hơn.

**GIỚI THIỆU ĐỀ TÀI**

1. Tổng quan về đề tài:

Đề tài dự đoán lượng truy cập website bằng phương pháp ARIMA được sử dụng trong bối cảnh đoán xem lượng truy cập trong các tháng tiếp theo là bao nhiêu.

1. Mô tả bài toán:

Chúng ta có một bộ dữ liệu chuỗi thời gian. Sau đó trải qua làm sạch và phân tích rồi áp dụng bộ dữ liệu vào mô hình ARIMA. Lựa chọn mô hình tốt nhất, đánh giá và đưa ra kết luận tổng quát cuối cùng.

**CHUỖI THỜI GIAN**

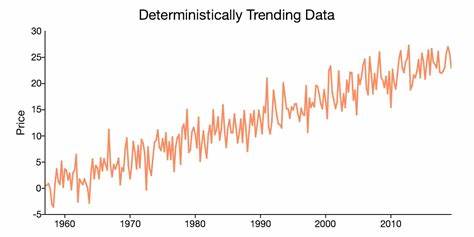
Chuỗi thời gian (time series) là một chuỗi các điểm dữ liệu được thu thập theo thứ tự thời gian. Mỗi điểm dữ liệu đại diện cho một giá trị của một đại lượng được quan sát tại một thời điểm cụ thể. Chuỗi thời gian được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau, chẳng hạn như kinh tế, tài chính, khoa học khí tượng, y tế và nhiều lĩnh vực khác.

Một số đặc điểm của chuỗi thời gian bao gồm:

* Xu hướng (Trend): Chuỗi thời gian có 1 xu hướng tăng hoặc giảm dài hạn
* Mùa vụ (Seasonality): Chuỗi thời gian có thể có một mô hình lặp lại theo chu kỳ cố định, chẳng hạn như hàng năm, hàng quý hoặc hàng tháng.
* Độ biến động (Volatility): Chuỗi thời gian có thể có độ biến động cao hoặc thấp, tức là giá trị của chuỗi thay đổi nhiều hoặc ít theo thời gian.
* T**ự tương quan (Autocorrelation):** Các giá trị trong chuỗi thời gian có thể tương quan với các giá trị trước đó của chính nó.

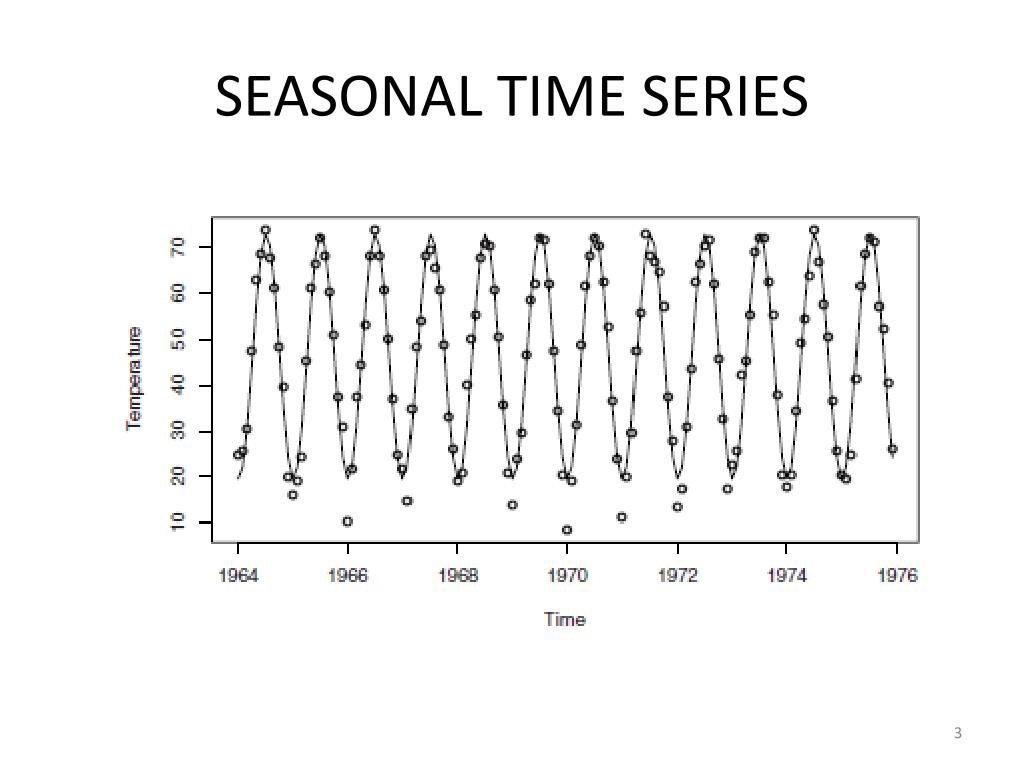
Các đặc điểm này có thể được phân tích và mô hình hóa để dự đoán các giá trị trong tương lai của chuỗi thời gian.

Chuỗi thời gian có xu hướng:



Ta có thể thấy chuỗi có đà tăng liên tục (có thể giảm) trong dài hạn

Chuỗi thời gian có tính mùa vụ:



Ta có thể thấy chuỗi thời gian lặp lại chu kỳ 1 năm. Cứ đạt đỉnh sau đó giảm rồi đạt đỉnh lần nữa sau 1 năm. Tương tự với đáy.

**MÔ HÌNH ARIMA**

1. Lý thuyết:

ARIMA là viết tắt của Autoregressive Intergrated Moving Average nghĩa là *tự hồi quy kết hợp trung bình động.* Đây là một mô hình phân tích thống kê sử dụng dữ liệu chuỗi thời gian để hiểu tốt hơn tập dữ liệu hoặc dự đoán xu hướng tương lai. Mô hình này có thể được hiểu bằng cách phân tích từng thành phần của nó như sau:

* 1. Auto regression:

Auto regression kí hiệu là AR. Đây là thành phần tự hồi quy dưa trên các giá trị bị trễ hoặc trước đó của nó. Độ trễ bậc ***p*** chính là giá trị lùi về quá khứ ***p*** bước thời gian của chuỗi. Độ trễ dài hoặc ngắn trong quá trình AR phụ thuộc vào tham số trễ ***p***. Cụ thể, quá trình AR(***p***) của chuỗi được biểu diễn như bên dưới:

AR(p) =

* 1. Moving average
  2. Interagated

1. Xây dựng mô hình:
2. Khảo sát: