

# **Hệ hỗ trợ quyết định dự báo doanh số bán hàng của siêu thị**

**Viện Toán ứng dụng và Tin học**  
**Đại học Bách khoa Hà Nội**

**Sinh viên thực hiện:** Nguyễn Đức Ánh  
**Giảng viên hướng dẫn:** TS. Lê Chí Ngọc

**Học phần:** Hệ hỗ trợ quyết định  
Hà Nội, tháng 6/2023



## Mục lục

- 1 Điều tra và khảo sát dự án
- 2 Đặc tả yêu cầu hệ thống
- 3 Thu thập dữ liệu
- 4 Thuật toán
- 5 Phân tích dữ liệu
- 6 Kết luận

## Mục lục

**1** Điều tra và khảo sát dự án

2 Đặc tả yêu cầu hệ thống

3 Thu thập dữ liệu

4 Thuật toán

5 Phân tích dữ liệu

6 Kết luận

## Điều tra và khảo sát

- Với nhu cầu tiêu dùng ngày càng tăng và sự cạnh tranh khốc liệt trên thị trường, các siêu thị có mong muốn biết được những gì phù hợp nhất với doanh nghiệp của họ.
- Các siêu thị muốn biết họ nên nhấn mục tiêu nào hoặc tránh những sản phẩm, khu vực, danh mục và phân khúc khách hàng nào, với mục đích đưa doanh nghiệp tăng trưởng.

## Đặt vấn đề và giải pháp

Đặt ra một số vấn đề cần thực hiện như sau

- ▣ Dữ liệu được đưa vào hệ thống, được thống kê và trực quan hóa như thế nào.
- ▣ Chức năng của hệ thống và những yêu cầu cần đảm bảo.
- ▣ Lựa chọn thuật toán để phân tích dữ liệu.

### Giải pháp

- ▣ Sử dụng thuật toán SARIMAX, áp dụng với chuỗi thời gian để dự báo doanh số bán hàng và doanh số đối với từng loại sản phẩm.
- ▣ Trực quan hóa dữ liệu một cách khái quát, sinh động, hỗ trợ nhà quản lý trong việc đưa ra quyết định.

## Mục lục

1 Điều tra và khảo sát dự án

2 Đặc tả yêu cầu hệ thống

3 Thu thập dữ liệu

4 Thuật toán

5 Phân tích dữ liệu

6 Kết luận

## Yêu cầu chức năng

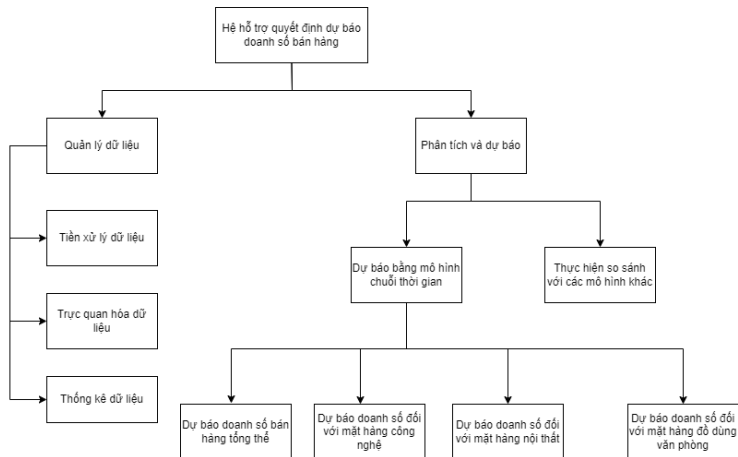
- ▣ **Dữ liệu:** được đưa lên hệ thống và được trực quan hóa.
- ▣ **Dự báo doanh số:** dự báo doanh số bán hàng của siêu thị, và dự báo doanh số bán hàng đối với các mặt hàng của siêu thị trong tương lai.

## Yêu cầu phi chức năng

- Độ tin cậy:  $MAPE < 30\%$ .
- Hệ thống: chạy trên nền tảng Web để người dùng có thể truy cập mọi lúc, mọi nơi.
- Có thể phục vụ nhiều người truy cập cùng lúc.
- Hệ thống có giao diện đơn giản, dễ sử dụng đối với nhà quản lý.



## Sơ đồ phân cấp chức năng



Hình: Sơ đồ phân cấp chức năng

## Mục lục

1 Điều tra và khảo sát dự án

2 Đặc tả yêu cầu hệ thống

3 Thu thập dữ liệu

4 Thuật toán

5 Phân tích dữ liệu

6 Kết luận

## Thu thập dữ liệu

Bộ dữ liệu thu thập được chứa các giao dịch của một siêu thị trong giai đoạn từ năm 2014 đến năm 2017 ở Mỹ, các giao dịch có thể được thực hiện xuyên giữa các bang ứng với các phương thức vận chuyển khác nhau. Các giao dịch được thực hiện với việc mua 3 loại hàng chính, trong đó được chia nhỏ thành 17 loại mặt hàng khác nhau.

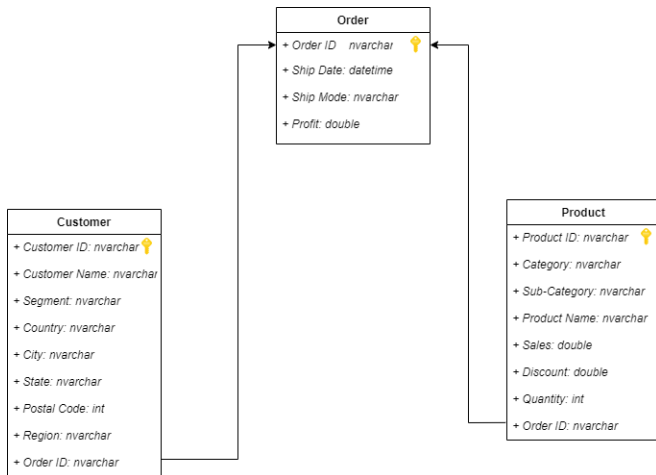
## Các trường dữ liệu

- ▣ Order ID: ID riêng đối với từng giao dịch.
- ▣ Ship Date: Thời gian đặt hàng.
- ▣ Ship Mode: Phương thức chuyển hàng.
- ▣ Customer ID: ID riêng đối với khách hàng.
- ▣ Customer Name: Tên khách hàng.
- ▣ Segment: Phân khúc mua.
- ▣ Country: Đất nước.
- ▣ City: Thành phố.
- ▣ State: Bang.
- ▣ Postal Code: Mã định danh.
- ▣ Region: Vùng.

## Các trường dữ liệu

- ▣ Product ID: ID sản phẩm.
- ▣ Category: Loại hàng.
- ▣ Sub-Category: Loại mặt hàng.
- ▣ Product Name: Tên sản phẩm.
- ▣ Sales: Giá bán.
- ▣ Quantity: Số lượng.
- ▣ Discount: Giảm giá.
- ▣ Profit: Lợi nhuận.

## Cơ sở dữ liệu quan hệ



Hình: Cơ sở dữ liệu quan hệ

## Mục lục

1 Điều tra và khảo sát dự án

2 Đặc tả yêu cầu hệ thống

3 Thu thập dữ liệu

**4 Thuật toán**

5 Phân tích dữ liệu

6 Kết luận

## Metric và tiêu chuẩn kiểm tra tính dừng

Metric được sử dụng ở đây là MAPE, được xác định bởi công thức sau

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i},$$

tại đó  $Y_i$ ,  $\hat{Y}_i$  lần lượt là giá trị thực và giá trị dự báo tại thời điểm  $i$ .

Trong bài trình bày, tác giả sử dụng tiêu chuẩn Dickey Fuller, được xét với cặp giả thuyết đối thuyết để kiểm tra tính dừng của chuỗi thời gian.



## Mô hình ARIMA

Mô hình ARIMA  $(p, q, d)$

$$\begin{aligned}A_t &= \nabla^d Y_t \\&= \gamma + \varepsilon_t + \phi_1 A_{t-1} + \phi_2 A_{t-2} + \dots + \phi_p A_{t-p} \\&\quad - v_1 Z_{t-1} - v_2 Z_{t-2} - \dots - v_q Z_{t-q}.\end{aligned}$$

$$\begin{aligned}f_{t+1} &= \gamma + \phi_1 A_t + \phi_2 A_{t-1} + \dots + \phi_p A_{t-p+1} \\&\quad - v_1 Z_t - v_2 Z_{t-1} - \dots - v_q Z_{t-q+1}.\end{aligned}$$

## Thuật toán SARIMAX

Thuật toán SARIMAX là cải tiến của thuật toán ARIMA, tại đó bên cạnh việc tích hợp giữa tự hồi quy (Autoregressive) và trung bình trượt (Moving Average), ta xét đến yếu tố mùa và các yếu tố ngoại sinh (Exogenous factors), giúp giảm sai số dự đoán, giải quyết vấn đề tự tương quan, và giúp cải thiện việc dự đoán <sup>1</sup>.

---

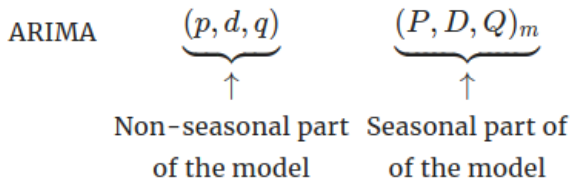
<sup>1</sup>Fahad Radhi Alharbi, Denes Csala, A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach

## Mô hình SARIMA

### Nhận xét

- Mô hình SARIMA về bản chất đây là mô hình ARIMA nhưng được điều chỉnh đặc biệt để áp dụng cho những chuỗi thời gian có yếu tố mùa vụ.
- Mô hình ARIMA bản chất là mô hình hồi quy tuyến tính nhưng mối quan hệ tuyến tính thường không giải thích tốt chuỗi trong trường hợp chuỗi xuất hiện yếu tố mùa vụ.

## Mô hình SARIMA



Hình: Mô hình SARIMA

## Mô hình SARIMAX

Mô hình toán học của thuật toán được phát biểu như sau

$$y_t = SARIMA(p, d, q)(P, D, Q)_m + \sum_{i=1}^m \beta_i X_i^t.$$

## Mục lục

1 Điều tra và khảo sát dự án

2 Đặc tả yêu cầu hệ thống

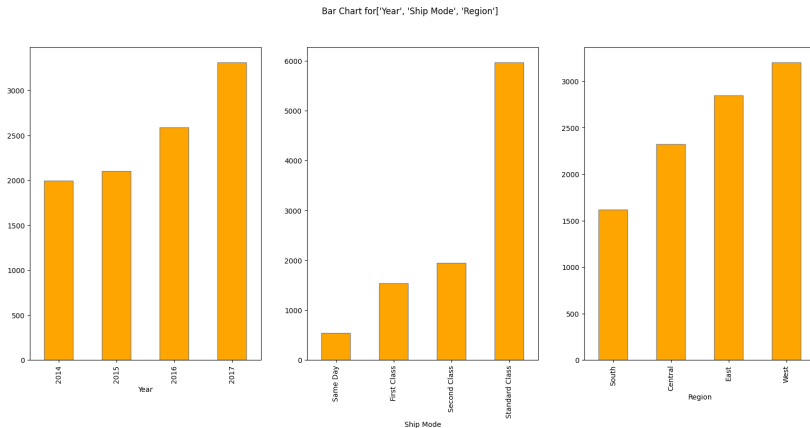
3 Thu thập dữ liệu

4 Thuật toán

5 Phân tích dữ liệu

6 Kết luận

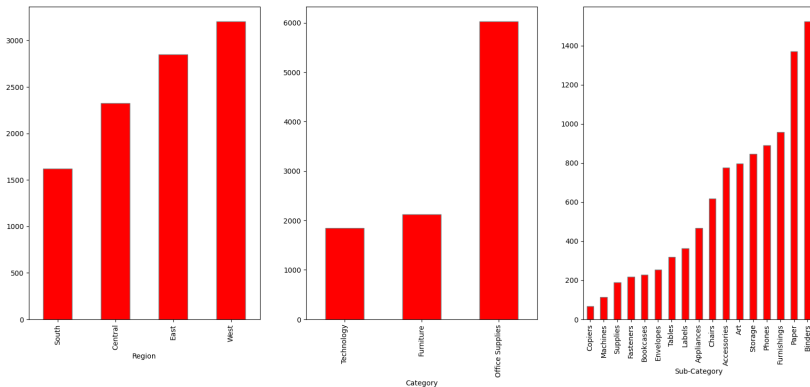
## Phân tích dữ liệu



Hình: Thống kê về năm, phương thức mua và vùng ứng với số giao dịch

## Phân tích dữ liệu

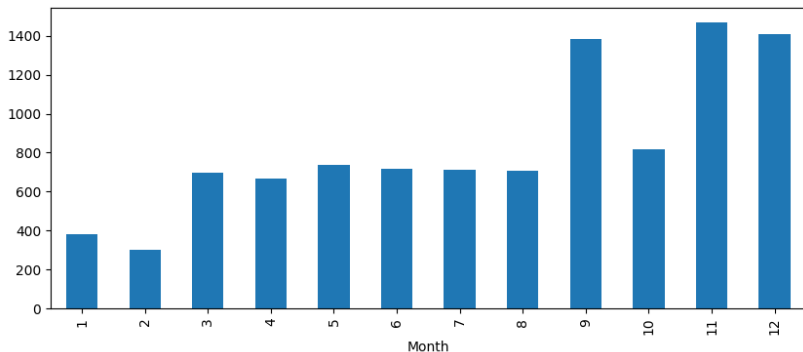
Bar Chart for ['Region', 'Category', 'Sub-Category']



Hình: Thống kê về vùng, loại hàng, loại mặt hàng ứng với số giao dịch

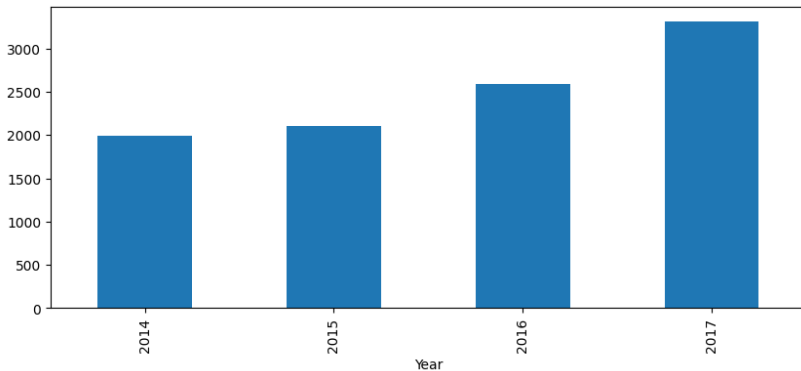


## Phân tích dữ liệu



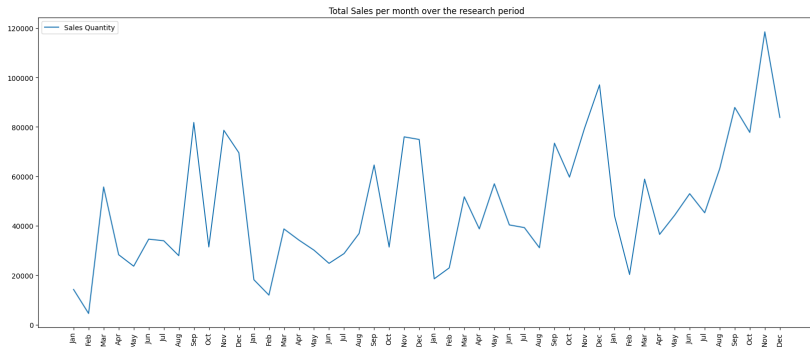
Hình: Thống kê theo tháng về số giao dịch

## Phân tích dữ liệu

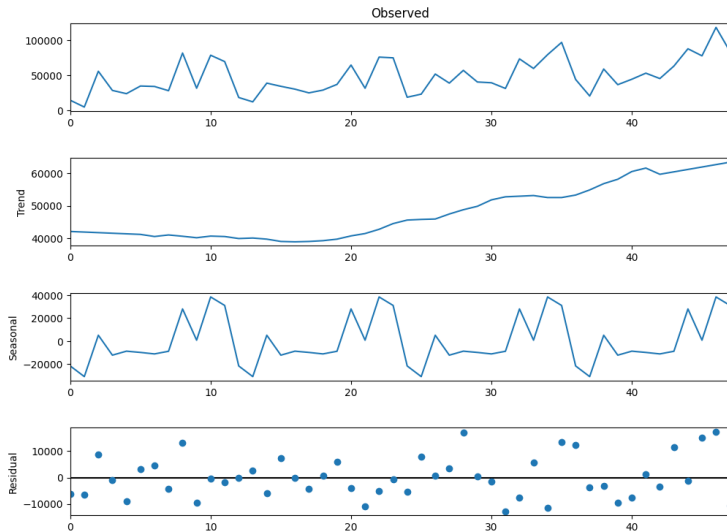


Hình: Thống kê theo năm về số giao dịch

## Phân tích dữ liệu

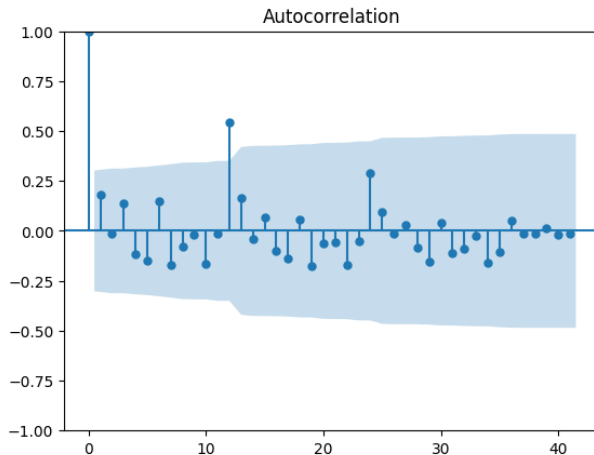


Hình: Tổng doanh số bán hàng theo từng tháng



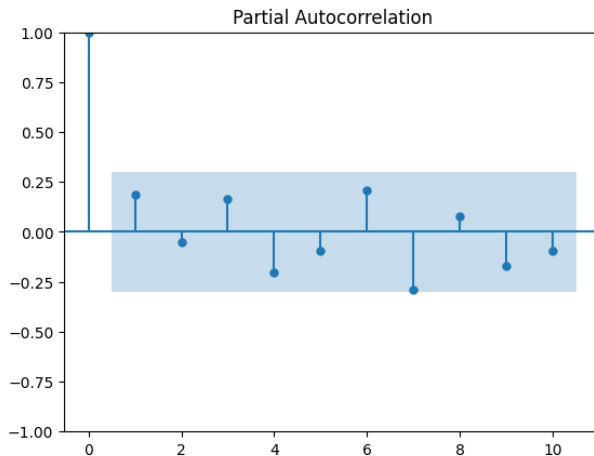
Hình: Mô hình cộng tính của chuỗi thời gian

## Đồ thị hàm tự tương quan (ACF)



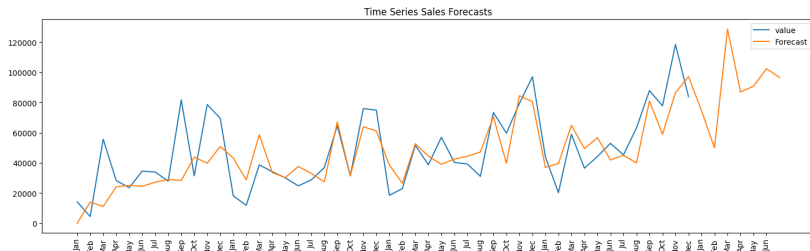
Hình: Đồ thị hàm tự tương quan

## Đồ thị hàm tự tương quan riêng (PACF)



Hình: Đồ thị hàm tự tương quan riêng

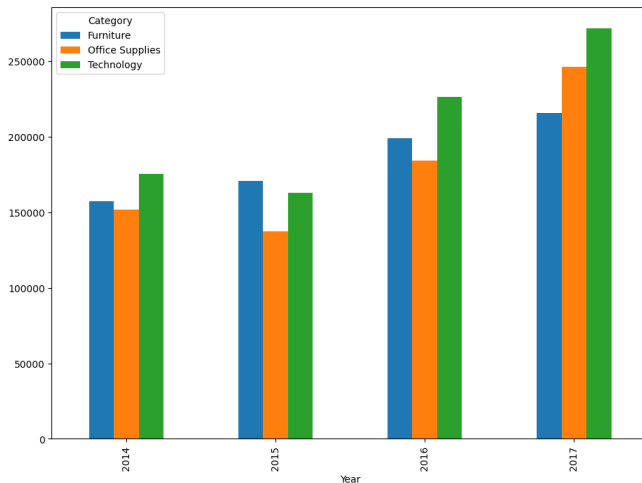
## Phân tích dữ liệu



Hình: Dự báo doanh số bán hàng trong 6 tháng tiếp theo

MAPE: 18.65%.

## Phân tích dữ liệu

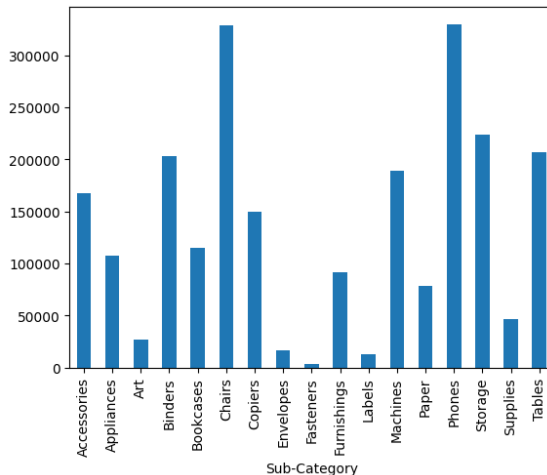


**Hình:** Tổng doanh số ứng với các loại hàng theo năm

Hệ hỗ trợ quyết định dự báo doanh số bán hàng của siêu thị



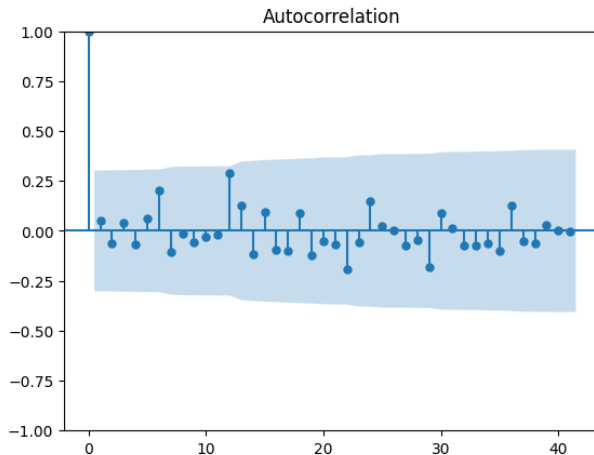
## Phân tích dữ liệu



**Hình:** Tổng doanh số ứng với các loại mặt hàng theo năm

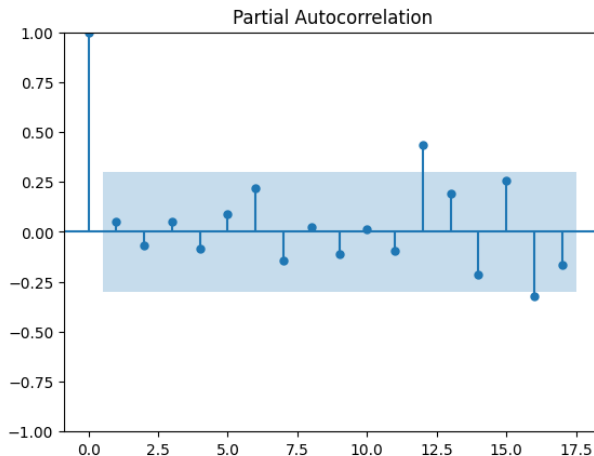
Hệ hỗ trợ quyết định dự báo doanh số bán hàng của siêu thị

## Đồ thị hàm tự tương quan (ACF)



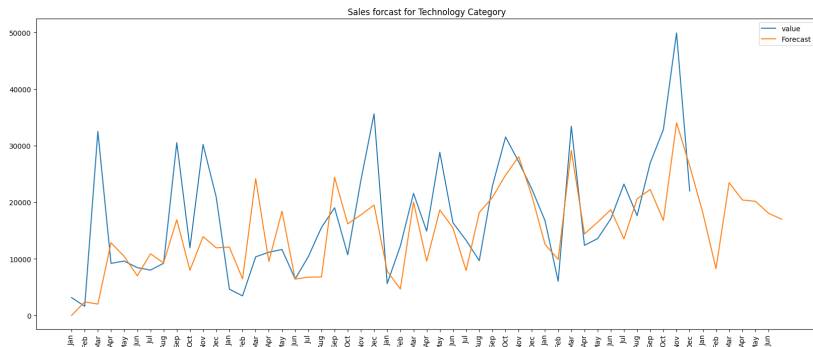
Hình: Đồ thị hàm tự tương quan với mặt hàng công nghệ

## Đồ thị hàm tự tương quan riêng (PACF)



Hình: Đồ thị hàm tự tương quan riêng với mặt hàng công nghệ

## Phân tích dữ liệu

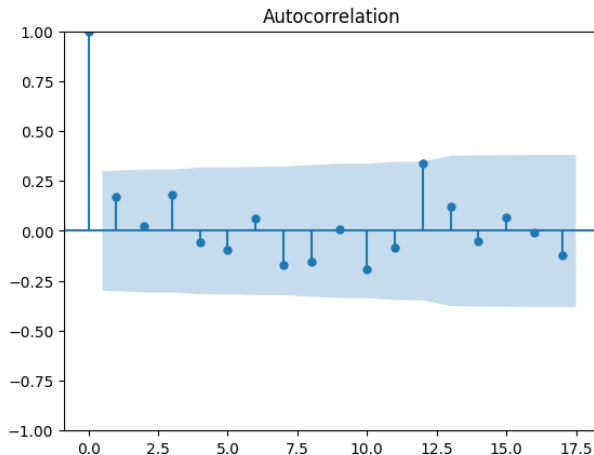


Hình: Dự báo doanh số bán hàng trong 6 tháng tiếp theo các mặt hàng công nghệ

MAPE: 27%

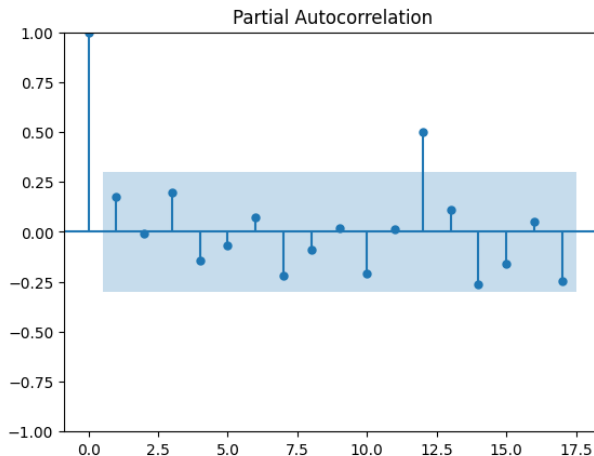
Hệ hỗ trợ quyết định dự báo doanh số bán hàng của siêu thị

## Đồ thị hàm tự tương quan (ACF)



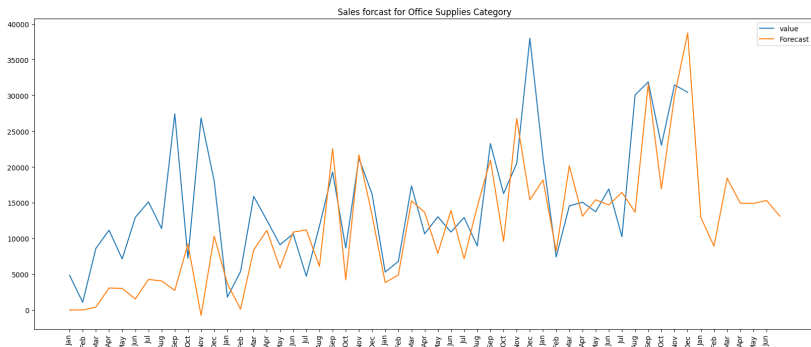
Hình: Đồ thị hàm tự tương quan với mặt hàng đồ dùng văn phòng

## Đồ thị hàm tự tương quan riêng (PACF)



Hình: Đồ thị hàm tự tương quan riêng với mặt hàng đồ dùng văn phòng

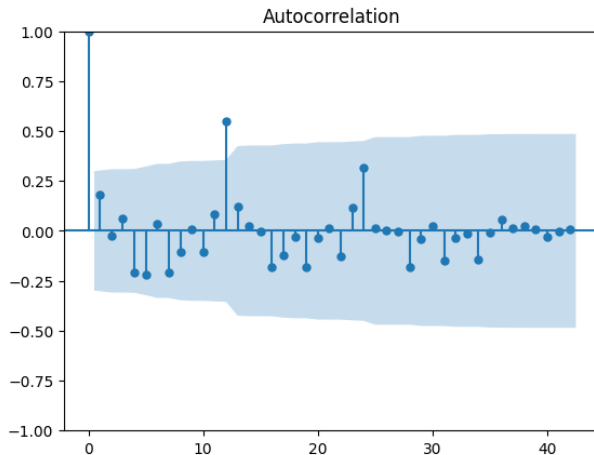
## Phân tích dữ liệu



Hình: Dự báo doanh số bán hàng trong 6 tháng tiếp theo các mặt hàng đồ dùng văn phòng

MAPE: 29.1%

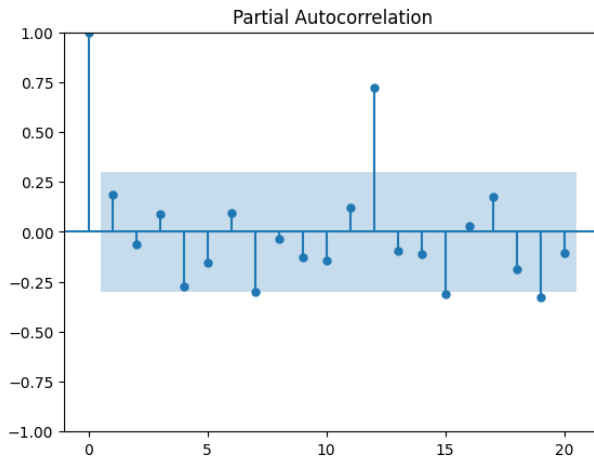
## Đồ thị hàm tự tương quan (ACF)



Hình: Đồ thị hàm tự tương quan ứng với Furniture

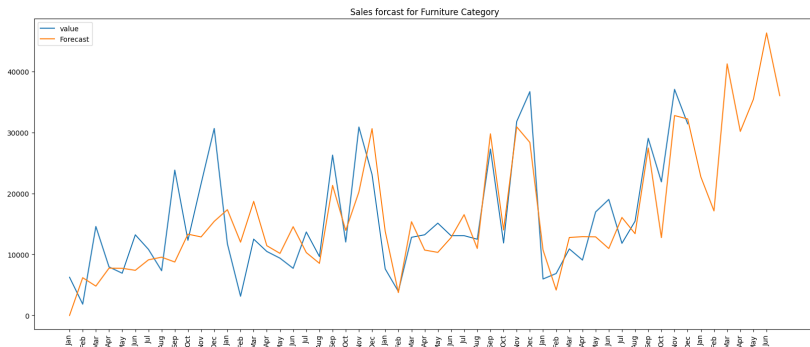


## Đồ thị hàm tự tương quan riêng (PACF)



Hình: Đồ thị hàm tự tương quan riêng ứng với Furniture

## Phân tích dữ liệu



Hình: Dự báo doanh số bán hàng trong 6 tháng tiếp theo các mặt hàng nội thất

MAPE: 14.92%.

## So sánh với các mô hình khác

Ta thực hiện so sánh MAPE của mô hình SARIMAX ứng với các mô hình Baseline gồm mô hình AR, MA, ARIMA.

Model Criteria	SARIMAX	ARIMA	AR	MA
Total Sales	18.65%	47.55%	38.85%	39.65%
Technology	27%	38.78%	37.65%	37.5%
Office Supplies	29.1%	54.54%	51.95%	54.52%
Furniture	14.92%	43.32%	41%	43.33%

**Bảng:** So sánh MAPE của các mô hình dự báo trên các khía cạnh khác nhau

## Mục lục

- 1 Điều tra và khảo sát dự án
- 2 Đặc tả yêu cầu hệ thống
- 3 Thu thập dữ liệu
- 4 Thuật toán
- 5 Phân tích dữ liệu
- 6 Kết luận

## Kết luận

Thông qua việc trực quan hóa dữ liệu và dự báo doanh số bán hàng bằng mô hình chuỗi thời gian, ta đưa ra một vài nhận định như sau

- Doanh số bán hàng đang tăng lên theo từng năm, trong đó các tháng đầu năm có xu hướng mua ít hơn, tăng dần về các tháng cuối năm.
- Các mặt hàng công nghệ có xu hướng được quan tâm nhiều càng về sau, cùng với sự tăng trưởng về tổng doanh số của mặt hàng đồ nội thất và đồ dùng văn phòng.
- Dựa theo dự báo này, nhà quản lý có thể tận dụng thời điểm để đưa sản phẩm vào thị trường, kịp thời nắm bắt nhằm tăng doanh thu và tạo ra sự tăng trưởng cho doanh nghiệp.

**Cảm ơn mọi người đã chú ý lắng nghe!**