

# Dự báo doanh thu bán hàng tập đoàn Vinamilk sử dụng mô hình ARIMA

Huỳnh Ngọc Quang

28/06/2024

## Tóm tắt nội dung

Bài báo cáo này trình bày phương pháp dự báo doanh thu bán hàng theo quý cho tập đoàn Vinamilk sử dụng mô hình ARIMA, một công cụ toán xác suất thống kê hiệu quả. Dữ liệu được thu thập từ các nguồn công khai và được xử lý bằng ngôn ngữ lập trình Python. Kết quả dự báo cho thấy mô hình ARIMA có thể cung cấp dự đoán chính xác với sai số chấp nhận được.

## 1 Lý do chọn đề tài và cơ sở lý thuyết

### 1.1 Lý do chọn đề tài

Dự báo doanh thu bán hàng là một trong những ứng dụng quan trọng của toán xác suất thống kê trong kinh doanh. Việc dự báo chính xác doanh thu giúp các doanh nghiệp lên kế hoạch sản xuất, tiếp thị, và quản lý tồn kho hiệu quả hơn.

### 1.2 Cơ sở lý thuyết

Mô hình ARIMA (Autoregressive Integrated Moving Average): Dựa trên giả thuyết chuỗi dừng và phương sai sai số không đổi. Mô hình sử dụng đầu vào chính là những tín hiệu quá khứ của chuỗi được dự báo để dự báo nó. Các tín hiệu đó bao gồm: chuỗi tự hồi qui AR (auto regression) và chuỗi trung bình trượt MA (moving average). Hầu hết các chuỗi thời gian sẽ có xu hướng tăng hoặc giảm theo thời gian, do đó yếu tố chuỗi dừng thường không đạt được. Trong trường hợp chuỗi không dừng thì ta sẽ cần biến đổi sang chuỗi dừng bằng sai phân. Khi đó tham số đặc trưng của mô hình sẽ có thêm thành phần bậc của sai phân  $d$  và mô hình được đặc tả bởi 3 tham số ARIMA( $p, d, q$ ).

Mô hình ARIMA là một trong những phương pháp thống kê phổ biến để dự báo chuỗi thời gian. ARIMA kết hợp ba thành phần chính:

- *AR* (AutoRegressive): Thành phần tự hồi quy bao gồm tập hợp các độ trễ của biến hiện tại
- *I* (Integrated): Quá trình đồng tích hợp hoặc lấy sai phân
- *MA* (Moving Average): Thành phần trung bình trượt

Mô hình ARIMA có thể nắm bắt được các đặc điểm như xu hướng và mùa vụ trong chuỗi thời gian, giúp cải thiện độ chính xác của dự báo.

## 2 Nội dung lý thuyết

### 2.1 Mô hình ARIMA

Mô hình ARIMA ( $p, d, q$ ) được xác định bởi ba tham số:

- $p$ : Số lượng các giá trị tự hồi quy
- $d$ : Số lần lấy sai phân để làm cho chuỗi thời gian trở nên dừng
- $q$ : Số lượng các giá trị trung bình trượt

Ngoài ra, mô hình SARIMA thêm các thành phần mùa vụ vào mô hình ARIMA thông qua các tham số mùa vụ ( $P, D, Q, m$ ).

### 2.2 Đánh giá mô hình

Mô hình ARIMA được đánh giá thông qua các chỉ số như AIC (Akaike Information Criterion) và BIC (Bayesian Information Criterion). Các chỉ số này cho biết mức độ phù hợp của mô hình với dữ liệu thực tế và giúp chọn ra mô hình tối ưu.

## 3 Phương pháp

### 3.1 Thu thập dữ liệu

Dữ liệu doanh thu bán hàng được thu thập từ các nguồn công khai, bao gồm các thông tin về doanh thu bán hàng thực tế theo quý.

### 3.2 Phân tích dữ liệu

Sử dụng Python, dữ liệu được xử lý và phân tích bằng thư viện 'pandas' và 'statsmodels'. Mô hình ARIMA được xây dựng và kiểm tra độ chính xác bằng cách chia dữ liệu thành tập huấn luyện và tập kiểm tra.

## 4 Kết quả

### 4.1 Trình bày kết quả

Kết quả phân tích cho thấy mô hình ARIMA có thể dự báo doanh thu bán hàng với độ chính xác chấp nhận được. Các chỉ số AIC và BIC được sử dụng để đánh giá mô hình.

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
import warnings
warnings.filterwarnings("ignore")
```

```
# Sample data
data = {
```

```

    'Quarter': ['2024Q1', '2023Q4', '2023Q3', '2023Q2', '2023Q1',
                '2022Q4', '2022Q3', '2022Q2', '2022Q1',
                '2021Q4', '2021Q3', '2021Q2', '2021Q1', '2020Q4',
                '2020Q3', '2020Q2', '2020Q1'],
    'Sales': [14112, 15619, 15637, 15195, 14000, 15069, 16094,
              14960, 13940, 15819, 16208, 15729, 13190,
              14425, 15563, 15495, 14150]
}

# Load data into DataFrame
df = pd.DataFrame(data)

# Convert Quarter to datetime

def convert_quarter_to_date(quarter_str):
    year, quarter = quarter_str[:4], quarter_str[4:]
    month = {'Q1': 1, 'Q2': 4, 'Q3': 7, 'Q4': 10}[quarter]
    return pd.Timestamp(f'{year}-{month:02d}-01')

df['Quarter'] = df['Quarter'].apply(convert_quarter_to_date)
df.set_index('Quarter', inplace=True)

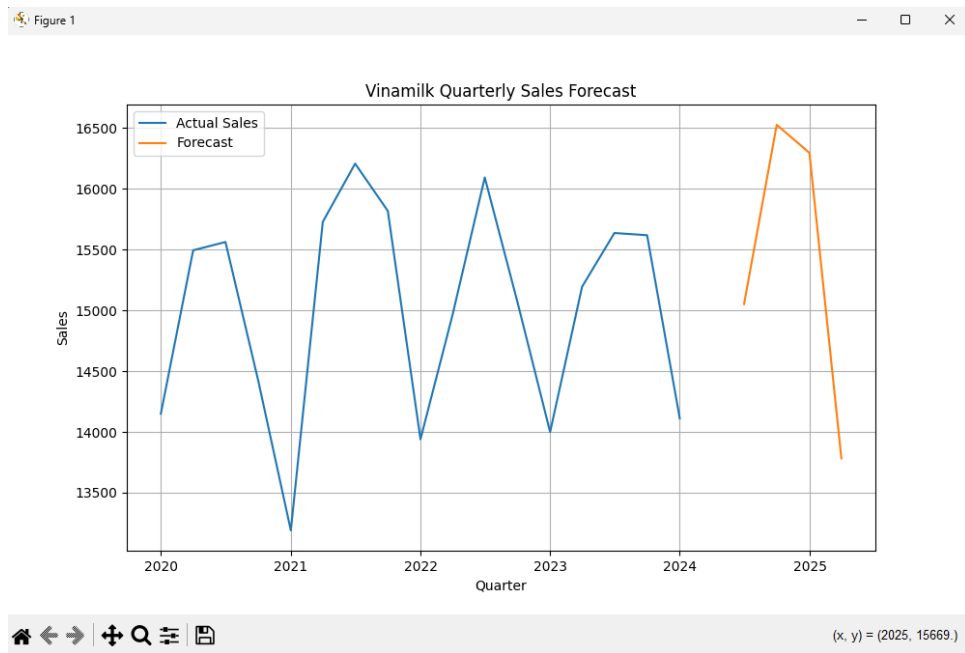
# Fit ARIMA model
model = ARIMA(df['Sales'], order=(7, 1, 2)) # seasonal_order
fit = model.fit()

print(fit.summary())

# Forecast
forecast_steps = 4 # Predict the next 4 quarters
forecast = fit.forecast(steps=forecast_steps)
forecast_dates = pd.date_range(
    start=df.index[0], periods=forecast_steps + 1, freq='Q')[1:]

# Plot results
plt.figure(figsize=(10, 6))
plt.plot(df.index, df['Sales'], label='Actual_Sales')
plt.plot(forecast_dates, forecast, label='Forecast', linestyle='
-')
plt.xlabel('Quarter')
plt.ylabel('Sales')
plt.title('Vinamilk_Quarterly_Sales_Forecast')
plt.legend()
plt.grid(True)
plt.savefig('forecast.png')
plt.show()

```



Hình 1: Biểu đồ ARIMA - Doanh thu theo quý

## 5 Kết luận

Biểu đồ trên thể hiện dự báo doanh số hàng quý của Vinamilk, với các điểm số liệu thực tế từ năm 2020 đến năm 2024, và dự báo cho năm 2025. Dưới đây là một số điểm đáng chú ý từ biểu đồ:

### 1. Xu hướng doanh số thực tế:

- Doanh số thực tế có sự dao động mạnh mẽ qua các quý.
- Các năm 2021 và 2023 có những đỉnh doanh số cao nhất, trong khi các quý khác doanh số giảm đáng kể.

### 2. Dự báo doanh số năm 2025:

- Biểu đồ thể hiện một sự tăng trưởng mạnh mẽ từ đầu năm 2025, sau đó giảm đáng kể vào các quý cuối năm.
- Sự dao động này có thể phản ánh các yếu tố mùa vụ hoặc những thay đổi dự kiến trong thị trường.

### 3. So sánh giữa thực tế và dự báo:

- Dự báo cho thấy xu hướng tăng trưởng ban đầu nhưng sau đó giảm, có thể cần xem xét lại các yếu tố đầu vào trong mô hình dự báo để đảm bảo tính chính xác.

### 4. Đánh giá chung:

- Biểu đồ cung cấp cái nhìn tổng quan về tình hình kinh doanh của Vinamilk và giúp dự đoán các xu hướng tương lai.

- Cần phân tích thêm về nguyên nhân của các dao động mạnh trong doanh số thực tế cũng như các yếu tố ảnh hưởng đến dự báo.

Mô hình ARIMA đã chứng minh tính hiệu quả trong việc dự báo doanh thu bán hàng dựa trên dữ liệu doanh thu các quý trước đó. Tuy nhiên, để nâng cao độ chính xác, có thể cần xem xét thêm nhiều biến số khác và các phương pháp dự báo phức tạp hơn trong tương lai.

## 6 Tài liệu tham khảo

- <https://phamdinhhkhanh.github.io/2019/12/12/ARIMAmode1.html>