

# Ứng dụng các mô hình chuỗi thời gian trong phân tích dự báo doanh thu theo ngày và tìm giá tối ưu của một cơ sở thuộc Công ty Minh Hà

Chuyên đề tốt nghiệp chuyên ngành Toán Kinh Tế

Sinh viên: Trương Ngọc Thùy Trang

Mã sinh viên : 11217004 - TOKT63

Giảng viên HD: ThS.Bùi Dương Hải

Hà Nội, Tháng 12-2024

- Dự báo doanh thu là hoạt động quan trọng, giúp doanh nghiệp lập kế hoạch tài chính và tối ưu hóa hoạt động kinh doanh.
- Việc đưa ra mức doanh thu mục tiêu trong lương lai cũng cần có căn cứ.
- Bên cạnh đó việc tìm mức giá bán để đạt được mục tiêu doanh thu hay lợi nhuận đề ra là cần thiết.

## Phương pháp nghiên cứu

- Sử dụng các mô hình chuỗi thời gian ARIMA, LSTM dự báo doanh thu.
- Tìm hàm cầu dựa vào mô hình VAR
- Tối ưu giá bán thông qua phương pháp tối ưu hoá và học tăng cường (Reinforcement Learning).

## Phạm vi nghiên cứu

- Dữ liệu từ cơ sở bán linh kiện của Công ty Minh Hà từ ngày 04/01/2021-30/01/2024
- Nghiên cứu tập trung vào chuỗi thời gian doanh thu, số lượng bán ra và giá bán.

- 1 Tổng quan vấn đề nghiên cứu
- 2 Dữ liệu - Thống kê mô tả
- 3 Mô hình và kết quả

- 1 Tổng quan vấn đề nghiên cứu
- 2 Dữ liệu - Thống kê mô tả
- 3 Mô hình và kết quả

# 1. Tổng quan vấn đề nghiên cứu

## Tổng quan thị trường linh kiện điện tử

- Thị trường linh kiện điện tử trên thế giới và tại Việt Nam đều đang trong giai đoạn phát triển mạnh mẽ, với nhiều cơ hội và thách thức.

## Giới thiệu Công ty Minh Hà

- Thành lập năm 2016, chuyên sản xuất và cung cấp linh kiện điện tử.
- Sản phẩm: IC, cảm biến, bo mạch,... và giải pháp công nghệ.
- Nhiều cơ sở hoạt động tại Hà Nội và Hồ Chí Minh.

# 1. Tổng quan vấn đề nghiên cứu

## Các nghiên cứu về dự báo doanh thu

- Soy Temür, A., Akgün, A., vs Temür, P. (2019). "Forecasting sales for real estate companies", Journal of Business Research, 98, 25-35.
- Hồ Khôi. (2020). "Xây dựng ứng dụng phân tích dự báo doanh thu doanh nghiệp Golf", Viện Hàn Lâm Khoa học và Công nghệ Việt Nam.

## Nghiên cứu về xác định hàm cầu

- Lutz Kilian và Cheolbeom Park (2009) nghiên cứu đề tài "The Demand for Energy and Price Volatility: A VAR Approach"

## Nghiên cứu về mô hình tối ưu giá

- Chandra et al. (2020) với bài nghiên cứu "AI Based Pricing Strategy in ECommerce"

- 1 Tổng quan vấn đề nghiên cứu
- 2 Dữ liệu - Thống kê mô tả**
- 3 Mô hình và kết quả



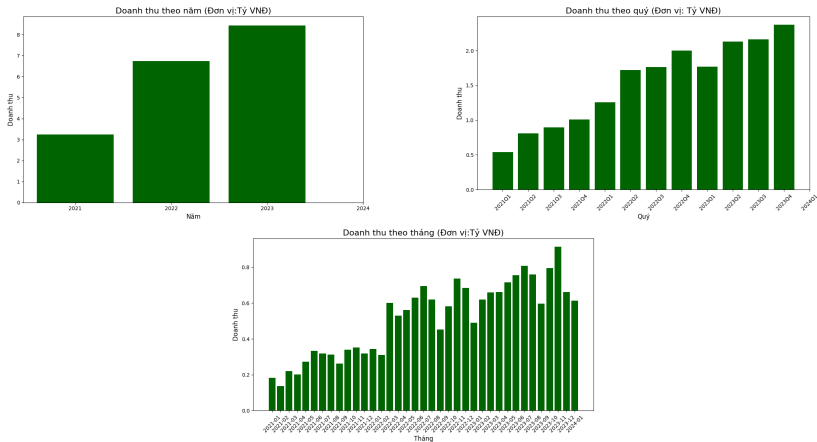
## 2. Dữ liệu - Thống kê mô tả

Công ty Minh Hà cung cấp bộ dữ liệu của một cơ sở bán lẻ điện tử thuộc công ty.

Bảng 1: Mô tả dữ liệu gốc

<b>Ngày (<math>t</math>)</b>	Từ ngày 04/01/2021 đến 30/01/2024 ( $t = 1, 2, \dots, 1066$ )
<b>Dòng sản phẩm (<math>i</math>)</b>	19 dòng sản phẩm thuộc Công ty Minh Hà ( $i = 1, 2, \dots, 19$ )
<b>Giá (<math>P_{it}</math>)</b>	Giá bán tại ngày $t$ của dòng sản phẩm $i$
<b>Số lượng (<math>Q_{it}</math>)</b>	Số lượng sản phẩm bán ra ngày $t$ của dòng sản phẩm $i$
<b>Doanh thu (<math>R_{it}</math>)</b>	Doanh thu ngày $t$ của dòng sản phẩm $i$

## 2.1. Phân tích doanh thu



Hình 1: Tổng doanh thu cơ sở theo năm, quý, tháng

## 2.2. Tính toán số liệu mới

Tổng doanh thu:  $TR_t = \sum_{i=1}^{19} R_{it}$

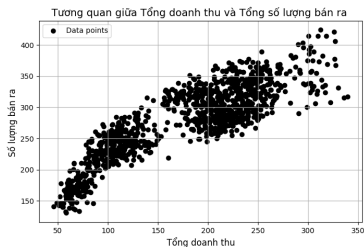
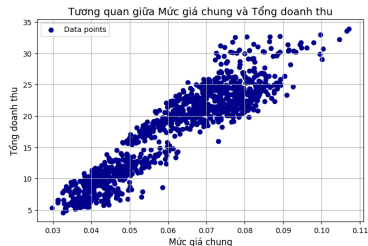
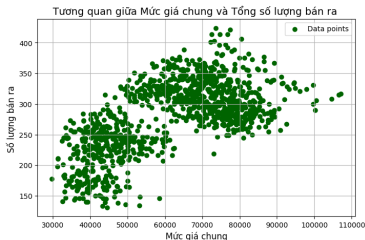
Tổng số lượng:  $TQ_t = \sum_{i=1}^{19} Q_{it}$

Để đảm bảo công thức  $R = P \cdot Q$ , nghiên cứu sẽ tính mức giá chung đại diện là trung bình có trọng số, với trọng số là tỷ lệ số lượng bán ra:

$$P_t^c = \frac{\sum_{i=1}^{19} P_{it} Q_{it}}{TQ_t} = \frac{\sum_{i=1}^{19} P_{it} Q_{it}}{\sum_{i=1}^{19} Q_{it}}$$

- $P_t^c$ : Mức giá chung của ngày  $t$ .
- $Q_{it}$ : Số lượng bán ra tại ngày  $t$  của sản phẩm  $i$ .
- $P_{it}$ : Giá bán tại ngày  $t$  của sản phẩm  $i$ .
- $R_{it}$ : Doanh thu tại ngày  $t$  của sản phẩm  $i$ .

## 2.3. Phân tích tương quan



Hình 2: Tương quan giữa các biến

- 1 Tổng quan vấn đề nghiên cứu
- 2 Dữ liệu - Thống kê mô tả
- 3 Mô hình và kết quả

### 3. Mô hình và kết quả

Nghiên cứu sẽ lần lượt thực hiện các bước sau đây:

- 1 Phân tích dự báo tổng doanh thu từ ARIMA và LSTM, chọn ra mô hình dự báo tối ưu nhất và coi đó là mức doanh thu công ty  $TR_t^*$  yêu cầu đạt được.
- 2 Tìm hàm cầu từ phân tích mô hình VAR giữa mức giá chung  $P_t^c$  và tổng số lượng  $TQ_t$ .
- 3 Tìm mức giá chung tối ưu thông qua phương pháp tối thiểu hoá hàm mục tiêu và phương pháp học tăng cường.

## 3.1. Mô hình ARIMA

ARIMA(0,1,1):

$$\Delta TR_t = -0.7115 \cdot \epsilon_{t-1} + \epsilon_t$$

ARIMA(5,1,1):

$$\begin{aligned}\Delta TR_t = & -0.1783\Delta TR_{t-1} - 0.0792\Delta TR_{t-2} - 0.0792\Delta TR_{t-3} \\ & -0.0669\Delta TR_{t-4} - 0.0373\Delta TR_{t-5} - 0.557\epsilon_{t-1} + \epsilon_t\end{aligned}$$

Trong đó:

- $\Delta TR_t$ : Sai phân bậc 1 của chuỗi  $TR_t$ .
- $\Delta TR_{t-k}$ : Sai phân bậc 1 tại độ trễ  $k$ .
- $\epsilon_t$ : Thành phần nhiễu trắng tại thời điểm  $t$ .
- $\epsilon_{t-1}$ : Thành phần nhiễu trắng tại thời điểm  $t - 1$ .

## 3.2. Mô hình LSTM

Để chuẩn bị cho mô hình, chia chuỗi tổng doanh thu thành tập huấn luyện (train) và tập kiểm thử (test) với tỷ lệ 80:20. Sau đó, chuẩn hoá dữ liệu chuỗi doanh thu theo phương pháp chuẩn hoá MinMax.

```
# Xây dựng mô hình LSTM
np.random.seed(42)
tf.random.set_seed(42)
random.seed(42)
model = Sequential([
    LSTM(261, return_sequences=True, input_shape=(sequence_length, 1)),
    Dropout(0.2), # Giảm overfitting
    LSTM(261, return_sequences=False),
    Dropout(0.2),
    Dense(1) # Dự đoán giá trị doanh thu tiếp theo
])

model.compile(optimizer='adam', loss='mean_squared_error')

# Huấn luyện mô hình
model.fit(X_train, y_train, epochs=50, batch_size=32, validation_data=(X_test, y_test), verbose=1)

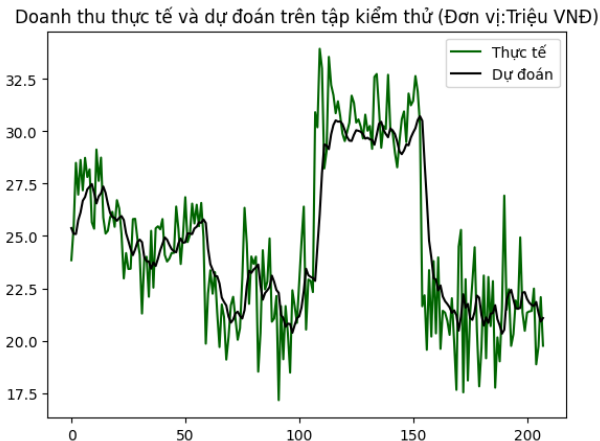
# Dự đoán
y_pred = model.predict(X_test)

# Chuyển dữ liệu về dạng gốc
y_test_original = scaler.inverse_transform(y_test)
y_pred_original = scaler.inverse_transform(y_pred)
```

Hình 3: Mô hình LSTM được xây dựng trong Python



## 3.2.Mô hình LSTM



Hình 4: Tổng doanh thu thực tế và dự đoán trên tập test

### 3.3.Đánh giá hiệu quả mô hình

Bảng 2: Đánh giá hiệu quả mô hình

Chỉ số	ARIMA(0,1,1)	ARIMA(5,1,1)	LSTM
RMSE	4015313.565	4014541.252	2277660.019
MAPE	13.581%	13.522%	6.936%

- Việc mô hình ARIMA(5,1,1) có RMSE và MAPE nhỏ hơn so với mô hình ARIMA(0,1,1) chứng tỏ rằng khi có đủ dữ liệu lịch sử và hiệu ứng tự hồi quy, mô hình phức tạp hơn có thể học hỏi và đưa ra dự đoán chính xác hơn.
- Mô hình LSTM có RMSE và MAPE đều nhỏ hơn đáng kể so với mô hình. Vậy nên mô hình LSTM sẽ tối ưu hơn trong việc dự báo doanh thu của cơ sở Công ty Minh Hà.
- Bài nghiên cứu này dùng kết quả dự báo doanh thu trong 30 ngày tiếp theo từ mô hình LSTM và xem đó là mức doanh thu mà Công ty yêu cầu đạt được trong 30 ngày này.

### 3.4. Hàm cầu từ mô hình VAR

- Kết quả kiểm định tính dừng cho ra sai phân bậc 1 của Mức giá chung và sai phân bậc 1 Tổng số lượng là chuỗi dừng với mức ý nghĩa lựa chọn 5%.
- Mô hình VAR cho hai chuỗi sai phân bậc 1 được lựa chọn theo tiêu chí AIC có độ trễ là 6.
- Khi đó hàm cầu có dạng:

$$\begin{aligned}TQ_t = & 0.389 - 0.00005P_{t-1}^c + 0.0002P_{t-2}^c - 0.00003P_{t-3}^c - 0.0002P_{t-4}^c \\& - 0.0001P_{t-5}^c + 0.0001P_{t-6}^c - 0.00003P_{t-7}^c + 0.361TQ_{t-1} + 0.183TQ_{t-2} \\& + 0.196TQ_{t-3} + 0.144TQ_{t-4} + 0.066TQ_{t-5} - 0.006TQ_{t-6} + 0.055TQ_{t-7}\end{aligned}$$

### 3.5. Tối ưu giá bán - Phương pháp tối thiểu hoá hàm mục tiêu

Hàm mục tiêu:

$$S = TR_t^* - TR_t$$

$$S = TR_t^* - P_t^c \cdot f(P_{t-1}^c, \dots, P_{t-7}^c, TQ_{t-1}, \dots, TQ_{t-7}) \rightarrow \min$$

Trong đó:

- $TR_t = P_t^c \cdot TQ_t$  : là hàm doanh thu
- $TQ_t = f(P_{t-1}^c, \dots, P_{t-7}^c, TQ_{t-1}, \dots, TQ_{t-7})$  : là hàm cầu được xác định từ mô hình VAR.
- $TR_t^*$  : Tổng doanh thu mục tiêu được xác định từ LSTM

### 3.5. Tối ưu giá bán - Phương pháp học tăng cường

- Học tăng cường (RL) là một phương pháp học máy mà trong đó, một tác nhân (agent) học cách đưa ra các quyết định tối ưu thông qua tương tác với môi trường và nhận phần thưởng (reward).
- Trong bài toán định giá, RL có thể được sử dụng để tìm giá bán tối ưu nhằm tối ưu hóa một mục tiêu, mục tiêu của nghiên cứu này tối thiểu hoá hàm S.

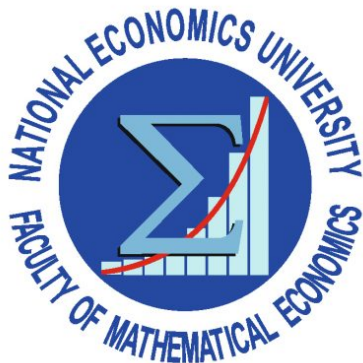
#### Quy trình tìm giá tối ưu

- 1 Tạo môi trường RL dựa trên dữ liệu lịch sử của công ty.
- 2 Q-Learning để tối ưu giá.
- 3 Huấn luyện mô hình RL.

### 3.5. Tối ưu giá bán - Kết quả

- Sau khi huấn luyện mô hình RL, mức giá tối ưu trung bình trong 30 ngày tiếp theo là 71161.487 VNĐ.
- So với mức giá tối ưu trung bình từ phương pháp tối thiểu hoá hàm mục tiêu là 70821.838 VNĐ thì mức giá tối ưu từ mô hình học tăng cường cao hơn 0.48%.
- Kết quả từ phương pháp tối thiểu hóa hàm mục tiêu có thể cung cấp một mức giá tối ưu trong điều kiện ổn định và biết trước các thông số như doanh thu mục tiêu và số lượng bán dự báo.
- Nhưng với học tăng cường, mức giá có thể được điều chỉnh theo thời gian, giúp tìm ra mức giá tối ưu hơn trong bối cảnh thay đổi và không chắc chắn.

- Sau quá trình nghiên cứu, chuyên đề đã tìm hiểu các phương pháp và mô hình liên quan đến phân tích và dự báo chuỗi thời gian.
- Ứng dụng các mô hình ARIMA và LSTM để dự báo doanh thu, tìm ra hàm cầu dựa vào mô hình VAR, đồng thời tìm giá bán tối ưu sản phẩm cho một cơ sở thuộc Công ty Cổ phần Công nghệ và Sản xuất Minh Hà.
- Việc kết hợp các mô hình mang đến một hướng tiếp cận mới mẻ. Khiến cho việc đưa ra chính sách về giá có căn cứ hơn.
- Tuy nhiên việc thử nghiệm vào thực tế để xem mức giá nào là mức giá tối ưu hơn thì cần thêm dữ liệu từ Công ty.



**Cảm ơn thầy cô đã lắng nghe!**