



BÁO CÁO VÒNG 1

ĐỀ BÀI: FORECASTING BUSINESS PERFORMANCE

ĐỘI THI: TND2004

Vũ Hải Đăng
Nguyễn Minh Nhật
Phạm Văn Thanh
Vũ Huyền Trâm



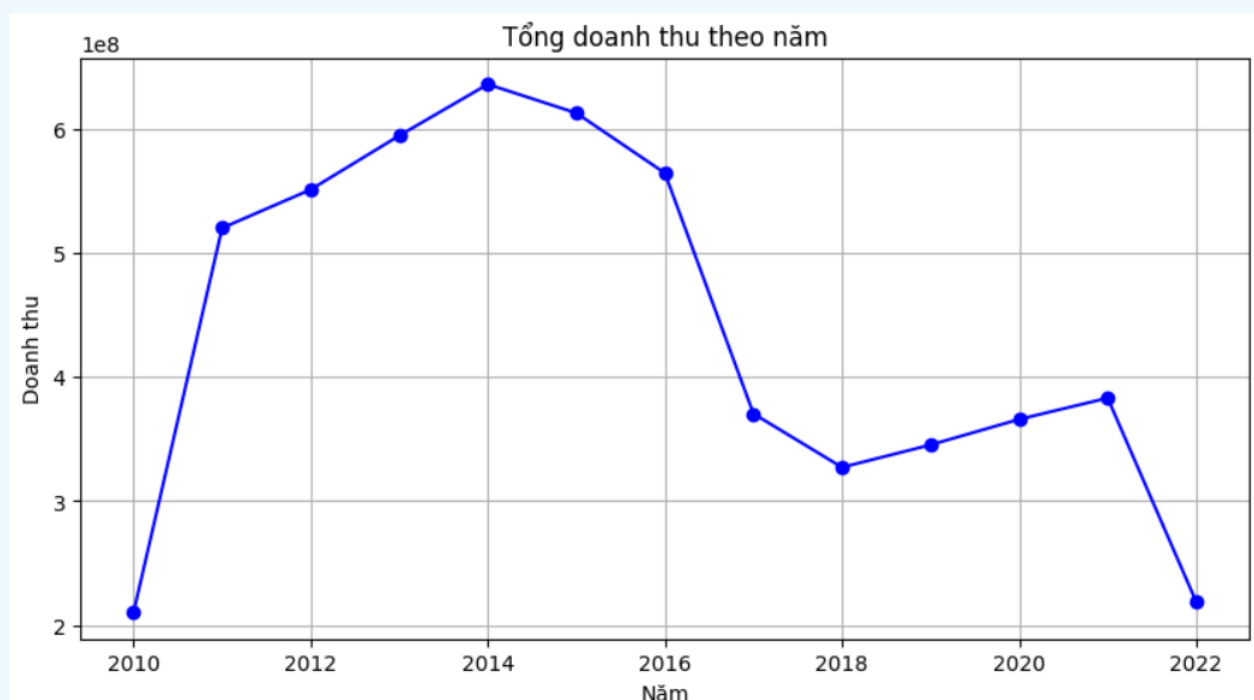
Mục lục

1	Phân tích dữ liệu	1
1.1	Sự thay đổi của doanh thu qua từng năm	1
1.2	Xu hướng của doanh thu trong một năm	2
1.2.1	Xu hướng theo quý	2
1.2.2	Xu hướng theo tháng	2
1.3	Xu hướng giá và sản lượng của các sản phẩm	3
1.4	Sự ảnh hưởng của yếu tố địa lý lên doanh thu	4
1.5	Tỷ trọng doanh thu của các phân khúc và loại mặt hàng	5
1.5.1	Tỷ trọng theo phân khúc (<i>Segment</i>)	5
1.5.2	Tỷ trọng theo loại hàng hóa (<i>Category</i>)	6
1.6	Sự ảnh hưởng của Units và COGS lên doanh thu	6
1.7	Tác động của yếu tố vĩ mô ảnh hưởng đến doanh thu	6
2	Mô hình dự đoán	7
2.1	Mô hình LSTM [1]	7
2.1.1	Xử lý đầu vào	7
2.1.2	Long Short-Term Memory (LSTM)	8
2.2	Mô hình Transformer	8
2.2.1	Đầu vào của mô hình	8
2.2.2	Xử lý Month_Index bằng Embedding	8
2.2.3	Kết hợp thông tin thời gian với doanh thu	9
2.2.4	Transformer Block để trích xuất đặc trưng	9
2.2.5	Dự báo doanh thu	9
2.2.6	Quá trình huấn luyện	9
2.3	Mô hình SARIMA	9
2.3.1	Đầu vào của mô hình	9
2.3.2	Kiểm định và xử lý tính dừng	10
2.3.3	Xác định tham số mô hình	10
2.3.4	Quá trình huấn luyện	10
2.3.5	Đánh giá mô hình	10
2.3.6	Phân tích phần dư	11
2.3.7	Dự báo doanh thu	11
3	Kết quả	11
3.1	Mô hình LSTM	11
3.2	Mô hình Transformer	12
3.3	Mô hình SARIMA	12
3.4	Đánh giá các mô hình	13
4	Dự đoán tình hình kinh doanh trong những năm tới	13
	References	13

1 Phân tích dữ liệu

1.1 Sự thay đổi của doanh thu qua từng năm

Quan sát xu hướng tổng thể, có thể thấy doanh thu của các bang tăng mạnh từ năm 2010 đến 2013, đạt đỉnh vào khoảng năm 2014 - 2015. Tuy nhiên, sau giai đoạn này, doanh thu có dấu hiệu suy giảm, đặc biệt từ năm 2016 đến 2019, với mức giảm khá rõ rệt. Giai đoạn 2020-2022, doanh thu có sự phục hồi nhẹ nhưng ngay lập tức suy giảm đột ngột. Điều này thường xảy ra với những mô hình kinh doanh có phân phối giữa các yếu tố tài chính (phân khúc, loại hàng hóa, yếu tố địa lý,...) không đồng đều, từ đó dẫn tới việc suy giảm một yếu tố chủ chốt nào đó có thể làm ảnh hưởng rõ rệt tới tổng doanh thu



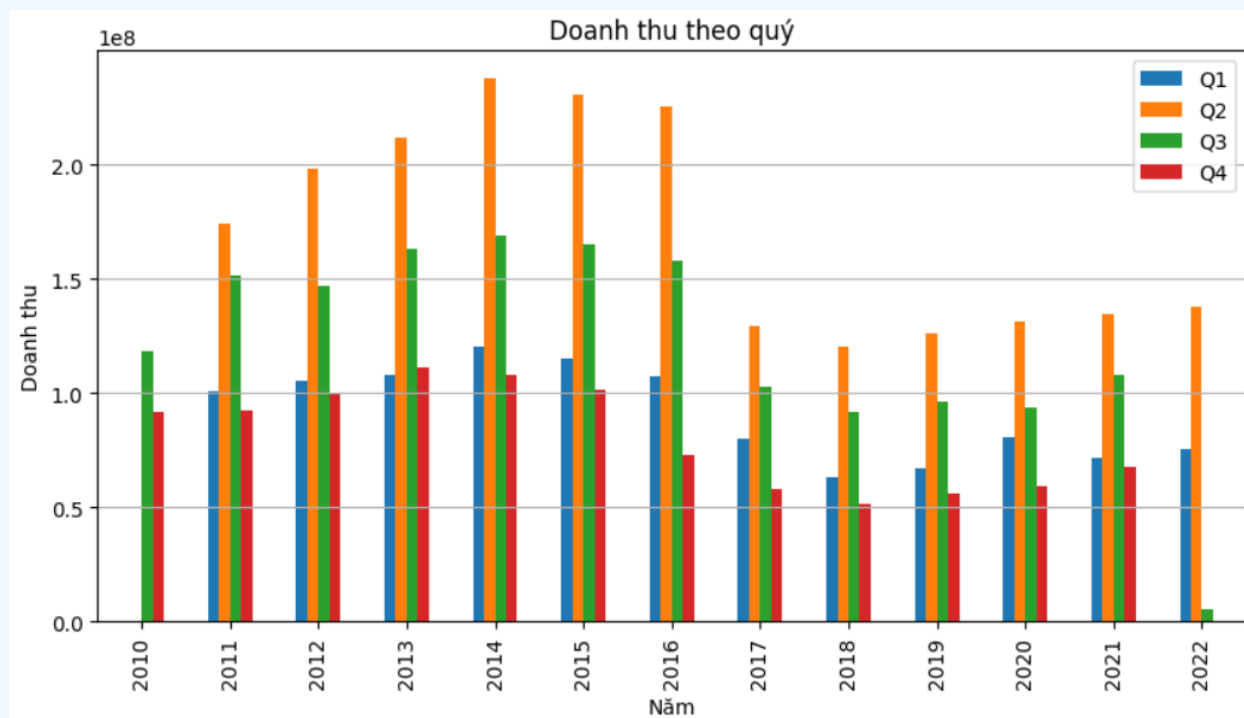
Hình 1: Biểu đồ tổng doanh thu theo từng năm

- **Năm 2010-2015:** Doanh thu tăng mạnh và đạt đỉnh vào khoảng 2014-2015.
- **Năm 2017-2018:** Doanh thu giảm mạnh, đánh dấu sự suy giảm đáng kể.
- **Năm 2019-2021:** Doanh thu duy trì ở mức thấp nhưng có sự phục hồi nhẹ.
- **Năm 2022:** Lại giảm mạnh, có thể do tác động từ các yếu tố bên ngoài như chính sách kinh tế, xã hội.

1.2 Xu hướng của doanh thu trong một năm

1.2.1 Xu hướng theo quý

Xu hướng biến động của doanh thu trong các quý không thay đổi nhiều qua các quý, qua đó cho thấy tính mùa vụ của doanh thu.



Hình 2: Biểu đồ doanh thu theo từng quý

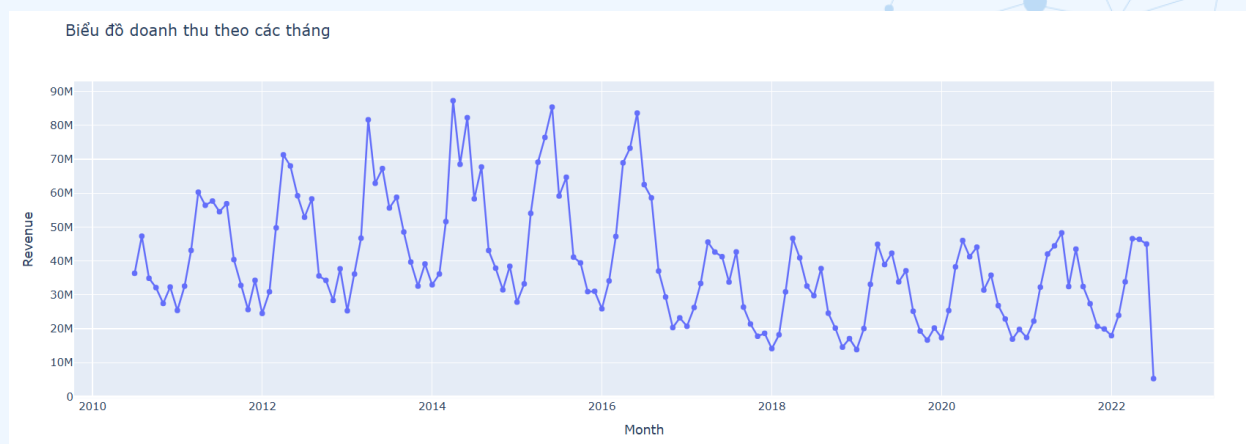
Quý 2 luôn có doanh thu cao nhất trong năm, cho thấy các sản phẩm của nhãn hiệu phù hợp với điều kiện thời tiết trong giai đoạn này. Kết quả của các quý sau cho thấy mức giảm rõ rệt, chạm đáy tại Quý 4 và có xu hướng hồi phục trở lại vào Quý 1 năm sau.

1.2.2 Xu hướng theo tháng

Đồ thị mô tả doanh thu theo tháng có xu hướng biến động tương tự với đồ thị theo quý:

- Doanh thu của hãng thường xuyên đạt đỉnh tại các tháng 4, 6.
- Suy giảm mạnh mẽ tại các tháng sau, chạm đáy vào tháng 12 và tháng 1 năm sau.

Điều này cho thấy doanh thu có thể bị ảnh hưởng bởi các yếu tố như điều kiện thời tiết, sự kiện đặc biệt hoặc hành vi tiêu dùng theo thời gian.

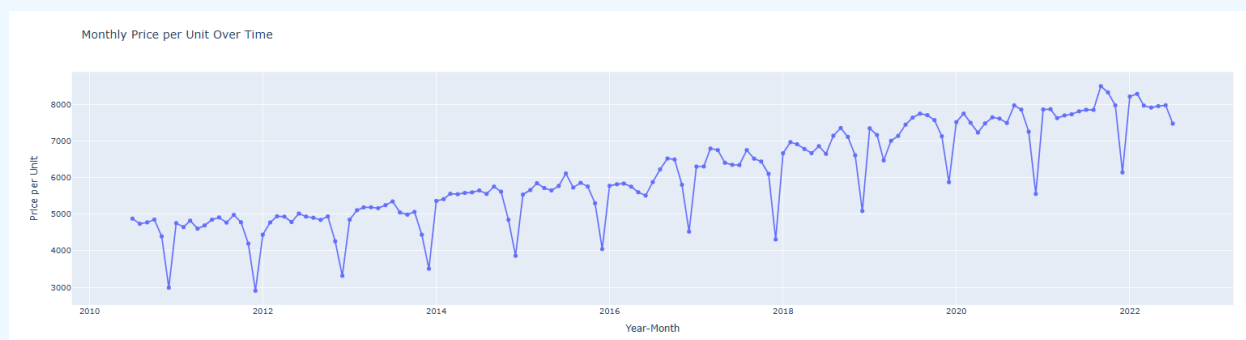


Hình 3: Biểu đồ doanh thu theo từng tháng

1.3 Xu hướng giá và sản lượng của các sản phẩm

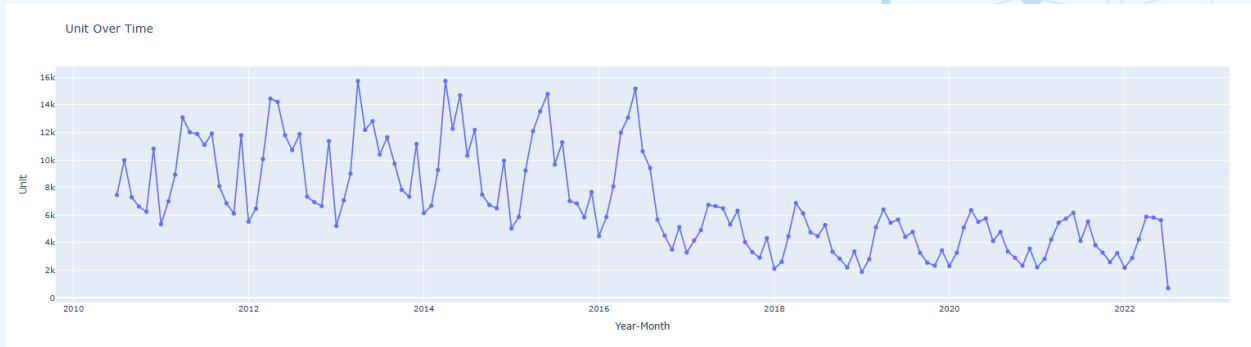
Biểu đồ cho thấy một xu hướng tăng đều đặn của giá bán trung bình trên mỗi đơn vị (*Price per Unit*) từ khoảng năm 2010 đến 2023. Mức giá ban đầu dao động quanh mốc 3.000–4.000, sau đó tăng dần lên gần 8.000 về cuối giai đoạn. Điều này cho thấy doanh nghiệp đang dần đẩy giá bán lên theo thời gian do lạm phát. **Tính mùa vụ và đáy vào tháng 12:**

Một điểm đáng chú ý là mỗi năm đều có một đợt giảm giá rõ rệt vào tháng 12, tạo ra đáy thấp nhất so với các tháng khác trong năm.



Hình 4: Biểu đồ giá trung bình sản phẩm được bán mỗi tháng

Đặc biệt, tháng cuối cùng (2022-07) ghi nhận mức giảm đáng kể với chỉ 705 đơn vị bán ra, thấp hơn hẳn so với các tháng trước đó. Sự sụt giảm đột ngột này khiến doanh thu giảm mạnh, đồng thời gây khó khăn cho các mô hình dự báo.

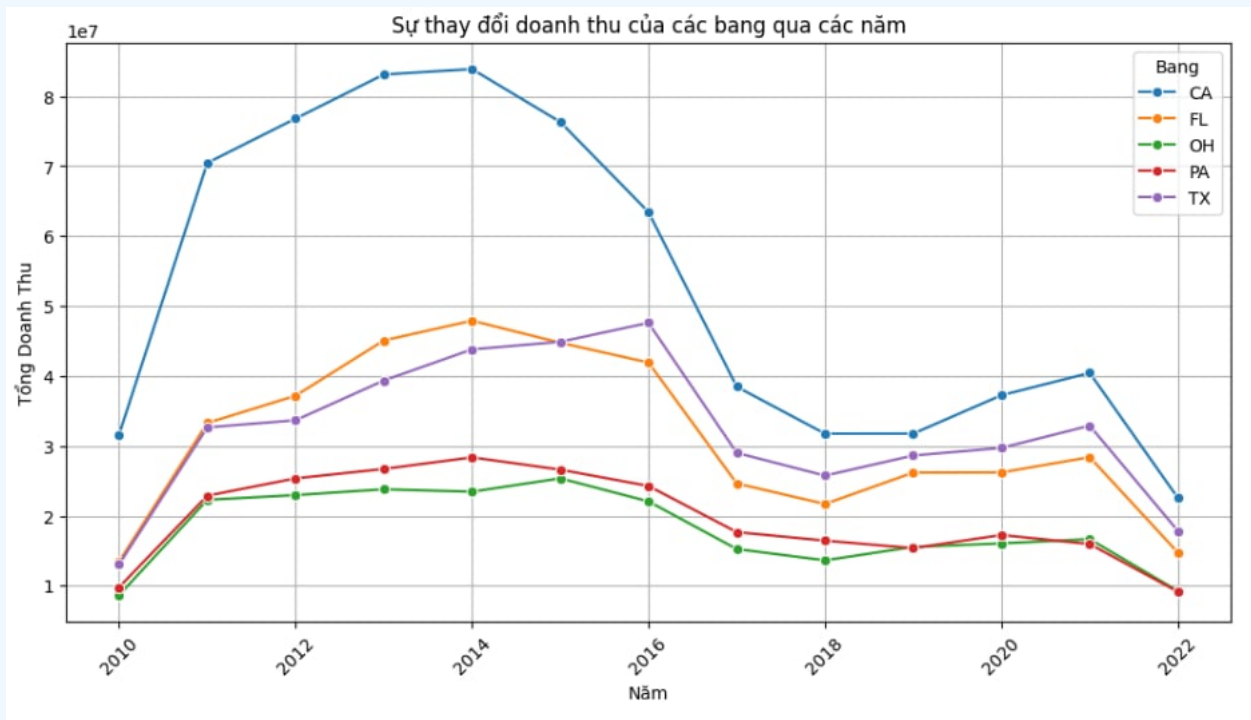


Hình 5: Số lượng sản phẩm được bán ra ở mỗi tháng qua các năm

1.4 Sự ảnh hưởng của yếu tố địa lý lên doanh thu

Một số bang có mức doanh thu vượt trội so với phần còn lại, điển hình là California (CA), với doanh thu đạt đỉnh hơn 80 triệu USD vào khoảng năm 2014 - 2015 trước khi suy giảm dần. Bên cạnh đó, các bang Florida (FL), New York (NY) và Texas (TX) cũng có doanh thu đáng kể, mặc dù vẫn thấp hơn so với California.

Trong khi đó, nhóm các bang có doanh thu trung bình duy trì mức tăng trưởng ổn định cho đến năm 2014 - 2016, sau đó có xu hướng giảm dần tương tự nhóm bang có doanh thu cao. Đối với các bang có doanh thu thấp hơn, xu hướng doanh thu ít biến động hơn nhưng vẫn chịu ảnh hưởng của các chu kỳ chung trên toàn quốc.



Hình 6: Sự thay đổi doanh thu của 5 bang có doanh thu lớn nhất

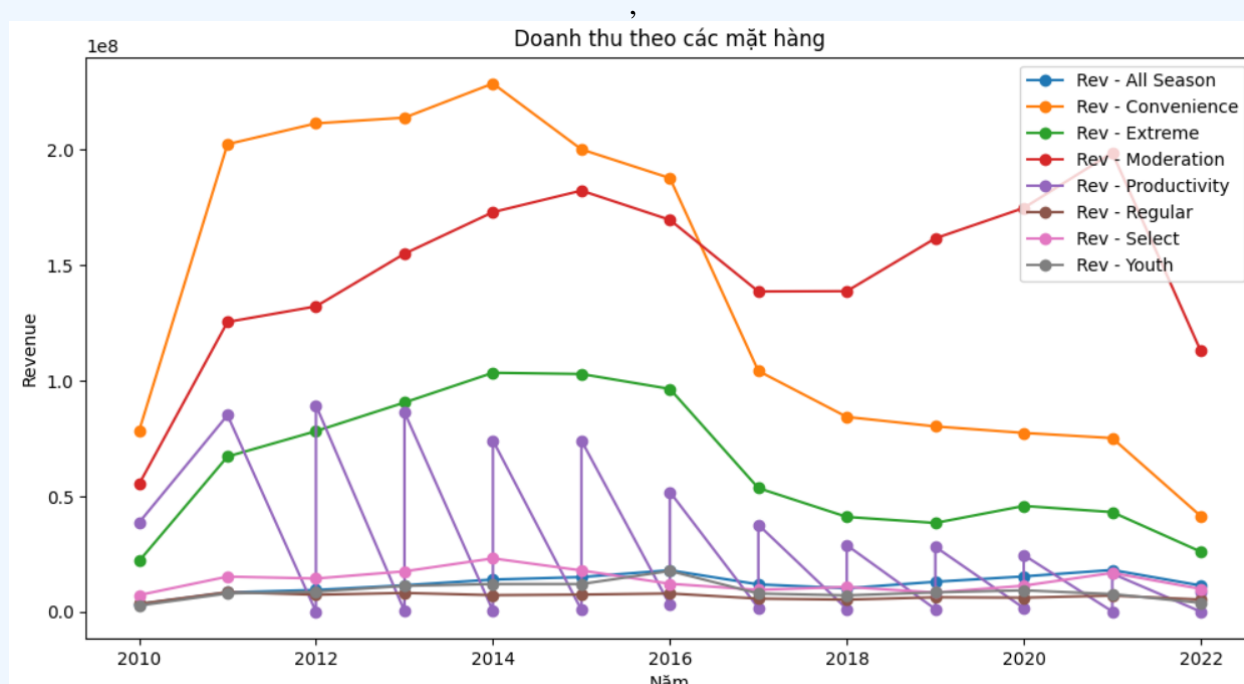
Forecasting Business Performance 1.5 Tỷ trọng doanh thu của các phân khúc và loại mặt hàng

Với xu hướng hiện tại, có thể dự đoán rằng một số bang lớn như California, Florida, New York và Texas sẽ tiếp tục đóng vai trò quan trọng trong tổng doanh thu.

1.5 Tỷ trọng doanh thu của các phân khúc và loại mặt hàng

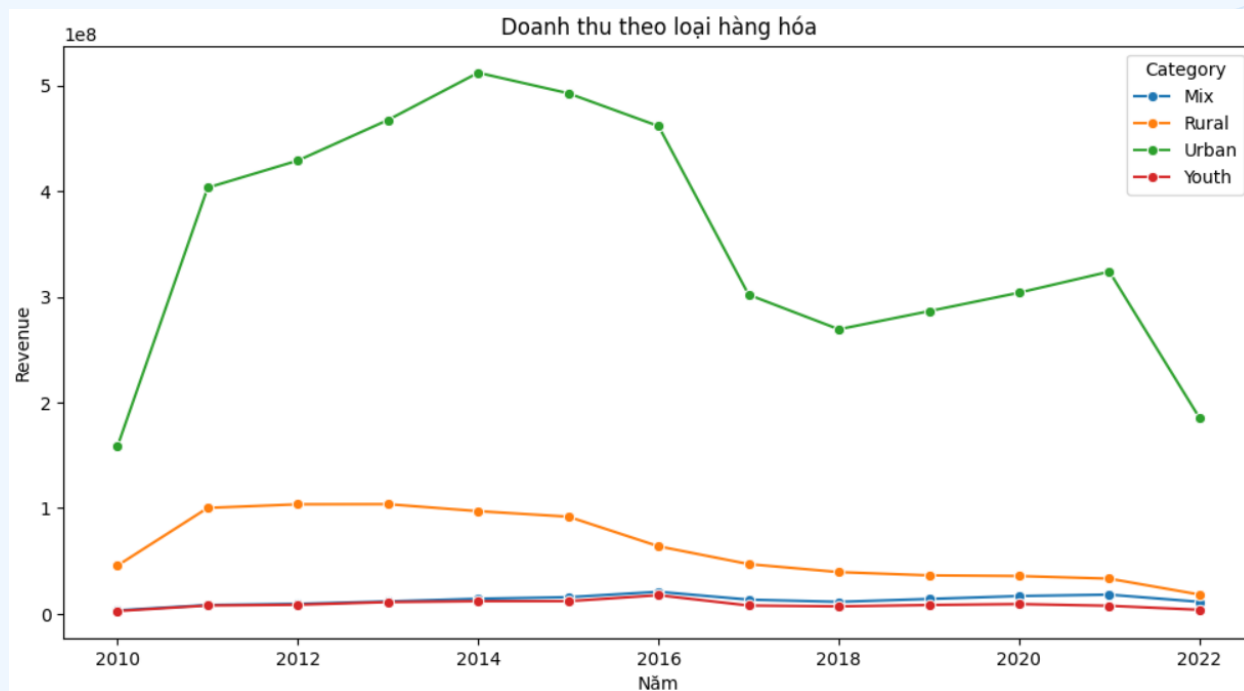
1.5.1 Tỷ trọng theo phân khúc (Segment)

Trong các dòng sản phẩm kinh doanh, có thể thấy mức doanh thu lớn thường đến từ các phân khúc *Convenience*, *Moderation*, *Extreme*. Tuy nhiên, xu thế chung của tất cả các dòng sản phẩm đều là giảm dần theo thời gian, điều này có thể suy ra tình hình kinh doanh trong những năm gần đây đang không hiệu quả



Hình 7: Doanh thu theo từng phân khúc

1.5.2 Tỷ trọng theo loại hàng hóa (Category)



Hình 8: Doanh thu theo các loại hàng hóa

Đối với các mặt hàng, ta có thể thấy sự chiếm đa số rõ ràng từ mặt hàng *Urban*. Doanh thu theo mặt hàng này có đường cong tương tự tổng doanh thu theo năm, đồng thời vượt trội hoàn toàn các mặt hàng khác. Có thể thấy rằng đây chính là nhân tố tác động chủ chốt tới doanh thu của nhân hàng này.

1.6 Sự ảnh hưởng của Units và COGS lên doanh thu

Doanh thu luôn cao hơn COGS, nhưng biên lợi nhuận có xu hướng giảm nhẹ, đòi hỏi chiến lược tối ưu hóa chi phí để duy trì hiệu quả kinh doanh.

Phân tích mô hình Random Forest cho thấy rằng COGS (Giá vốn hàng bán) có ảnh hưởng gần như tuyệt đối đến doanh thu, chiếm 98.86%, trong khi số lượng sản phẩm bán ra (Units) chỉ đóng góp 1.14%. Điều này cho thấy doanh thu phụ thuộc chủ yếu vào chi phí hàng hóa hơn là số lượng bán ra, có thể do sự biến động của giá bán theo từng giai đoạn. Vì vậy, để tối ưu doanh thu, doanh nghiệp nên tập trung vào quản lý COGS, tối ưu chi phí sản xuất và tìm nguồn cung hợp lý hơn, thay vì chỉ cố gắng tăng số lượng bán. Ngoài ra, cần cân nhắc thêm các yếu tố khác như chiến lược định giá, khuyến mãi hoặc xu hướng thị trường để có cách tiếp cận hiệu quả hơn.

1.7 Tác động của yếu tố vĩ mô ảnh hưởng đến doanh thu

Dựa trên phân tích dữ liệu, phân khúc Urban - bao gồm ba nhóm chủ đạo là Moderation, Convenience và Extreme - thể hiện xu hướng biến động đáng chú ý qua các năm. Trong giai đoạn

trước năm 2017, phân khúc này chiếm tỷ trọng lớn trong tổng doanh thu, với California - trung tâm của ngành thời trang đô thị - có đường doanh thu song hành cùng xu hướng tổng thể. Thị trường thời trang Urban tại California duy trì được sức hấp dẫn mạnh mẽ, bất chấp những ảnh hưởng từ các yếu tố vĩ mô như chính sách thuế và hạn chế nhập cư dưới thời Trump. Tuy nhiên, bước ngoặt đáng kể xuất hiện từ năm 2017, khi doanh thu sụt giảm mạnh xuống chỉ còn khoảng 50% so với giai đoạn đỉnh cao trước đó. Xu hướng này tiếp tục chịu thêm áp lực từ đại dịch COVID-19 và lạm phát trong những năm sau đó. Đặc biệt, sự sụt giảm đạt đến đỉnh điểm vào tháng 7/2022 - thời điểm được đánh dấu bởi nhiều yếu tố bất lợi cùng hội tụ. Lạm phát tại Mỹ chạm mốc kỷ lục 9.1% - cao nhất trong 40 năm, cùng với khủng hoảng năng lượng do xung đột Nga-Ukraine và hậu quả kéo dài của đứt gãy chuỗi cung ứng sau đại dịch đã tạo ra một "cơn bão hoàn hảo". Thêm vào đó, sự thay đổi mạnh mẽ trong thị hiếu người tiêu dùng, với xu hướng ưa chuộng trang phục casual và thoải mái hậu COVID-19, cũng góp phần làm suy giảm doanh số của các dòng sản phẩm Urban truyền thống.

Mặc dù các nhóm Moderation, Convenience và Extreme vẫn duy trì được một mức độ hấp dẫn nhất định, những biến động này phản ánh rõ sự thay đổi sâu sắc trong hành vi tiêu dùng. Người tiêu dùng ngày càng chuyển hướng sang mua sắm trực tuyến và quan tâm nhiều hơn đến thời trang bền vững, đòi hỏi các doanh nghiệp phải nhanh chóng thích ứng với những thay đổi này để khắc phục xu hướng tiêu cực trong doanh thu. Những biến động này không chỉ thể hiện tác động của các yếu tố vĩ mô lên thị trường, mà còn cho thấy sự cần thiết phải tái cấu trúc chiến lược kinh doanh trong phân khúc thời trang đô thị.

2 Mô hình dự đoán

Ở bài toán này do tính mùa vụ của dữ liệu, chúng tôi đề xuất và thực hiện ba mô hình LSTM, SARIMA và Transformer.

2.1 Mô hình LSTM [1]

2.1.1 Xử lý đầu vào

Mô hình được xây dựng với hai đầu vào chính: **Month_Index** và **Revenue**. Trong đó, **Month_Index** biểu diễn thông tin thời gian theo tháng, giúp mô hình nhận diện các yếu tố mùa vụ ảnh hưởng đến doanh thu. **Revenue** là chuỗi doanh thu lịch sử, đóng vai trò quan trọng trong việc học xu hướng biến động theo thời gian. Cụ thể, mô hình sử dụng dữ liệu doanh thu của 12 tháng trước để dự đoán doanh thu của tháng kế tiếp. Trước khi đưa vào mô hình, chuỗi thời gian này được chuẩn hóa bằng Min-Max Scaler về khoảng $[-1, 1]$ nhằm cải thiện hiệu suất học máy

$$input_t = embedding(month(t)) + pre_revenue(t, 12)$$

Với t là thời điểm (*Year-Month*) của dữ liệu đầu vào, $embedding(month(t))$ là vector nhúng của tháng tại thời điểm t , dùng để đặc trưng cho tính mùa vụ. $pre_revenue(t, 12)$ là kết quả kinh doanh của 12 tháng trước thời điểm t .

2.1.2 Long Short-Term Memory (LSTM)

Tín hiệu đầu vào sau khi được kết hợp sẽ được đưa vào một lớp **LSTM** được thiết lập với trạng thái duy trì (**stateful=True**), tức là mô hình có thể nhớ trạng thái từ các bước thời gian trước đó, giúp học tốt hơn các xu hướng dài hạn trong dữ liệu doanh thu.

Mô hình được huấn luyện với batch size bằng 1, được tối ưu hóa bằng thuật toán **Adam** và sử dụng hàm mất mát **custom_loss** để đo lường sai số dự báo.

Để đảm bảo mô hình có thể phù hợp với kết quả kinh doanh, chúng tôi tùy chỉnh hàm mất mát, mục tiêu giúp mô hình có thể đảm bảo thu được giá trị tốt với nhiều metric đánh giá.

Hàm loss tùy chỉnh **custom_loss** được định nghĩa dưới dạng tổng có trọng số của sai số bình phương trung bình (MSE) và sai số phần trăm tuyệt đối đối xứng (SMAPE), cụ thể:

$$custom_loss = MSE + \lambda \times SMAPE$$

Ở đây, MSE đo lường sai số tuyệt đối giữa giá trị dự báo và giá trị thực, trong khi SMAPE đánh giá sai số theo tỷ lệ phần trăm một cách đối xứng. Việc kết hợp này giúp mô hình tối ưu hóa cả sai số tuyệt đối lẫn tương đối, từ đó cải thiện độ chính xác trong dự báo.

Kết quả quá trình huấn luyện và validation cho thấy bộ tham số sau là tối ưu:

Neurons	Epochs	Learning_rate	λ
64	150	0.004	0.2

Bảng 1: Siêu tham số cho mô hình LSTM

2.2 Mô hình Transformer

2.2.1 Đầu vào của mô hình

Mô hình được thiết kế với hai đầu vào chính:

- **Month_Index**: Đại diện cho thông tin thời gian theo tháng, giúp mô hình học được yếu tố mùa vụ trong dữ liệu doanh thu.
- **Revenue**: Chuỗi doanh thu lịch sử, cung cấp thông tin quan trọng để mô hình dự báo xu hướng trong tương lai. Ở đây, chúng tôi sử dụng chuỗi dữ liệu doanh thu của 12 tháng trước để dự đoán doanh thu tháng tiếp theo.[\[2\]](#)

2.2.2 Xử lý Month_Index bằng Embedding

Tương tự, như mô hình LSTM, chỉ số tháng **Month_Index** được đưa vào một lớp **Embedding** với kích thước đầu ra là 12, giúp ánh xạ giá trị rời rạc của tháng thành một không gian liên tục, từ đó mô hình có thể học được sự tương quan giữa các tháng trong năm. Sau đó, một lớp **Layer**

Normalization được áp dụng để chuẩn hóa dữ liệu nhúng, giúp ổn định quá trình huấn luyện và tăng tốc độ hội tụ.

2.2.3 Kết hợp thông tin thời gian với doanh thu

Sau khi **Month_Index** được biến đổi thành vector nhúng, nó được **cộng trực tiếp** với dữ liệu **Revenue**. Việc kết hợp này cho phép mô hình sử dụng cả thông tin về mùa vụ và xu hướng doanh thu trước khi đưa vào bộ khối Transformer.

2.2.4 Transformer Block để trích xuất đặc trưng

Sau bước kết hợp dữ liệu, tín hiệu đầu vào sẽ được đưa vào một khối Transformer để trích xuất đặc trưng của chuỗi thời gian:

- **Multi-Head Attention:** Áp dụng **Self-Attention** với 8 đầu (heads), giúp mô hình tập trung vào các khoảng thời gian quan trọng trong chuỗi doanh thu.[3]
- **Residual Connection và Layer Normalization:** Đầu ra của khối Attention được cộng với đầu vào gốc (Residual Connection) và chuẩn hóa bằng **Layer Normalization**.
- **Feed-Forward Network (FFN):** Sử dụng hai lớp **Dense** với hàm kích hoạt **GELU** để học các quan hệ phi tuyến giữa các đặc trưng trích xuất từ Self-Attention.
- **Dropout:** Được áp dụng trong cả Multi-Head Attention và FFN để giảm hiện tượng overfitting.

2.2.5 Dự báo doanh thu

Đầu ra từ Transformer Block đưa vào một lớp **Dense** để tạo ra dự báo doanh thu tiếp theo.

2.2.6 Quá trình huấn luyện

Mô hình được huấn luyện trong **500 epochs** với batch size bằng 1. Learning rate được thiết lập là 0.001, sử dụng thuật toán **Adam** để tối ưu hóa mô hình, đồng thời sử dụng hàm mất mát **custom_loss** tương tự mô hình LSTM để đo lường sai số của dự báo.

2.3 Mô hình SARIMA

2.3.1 Đầu vào của mô hình

Mô hình SARIMA được xây dựng trên cơ sở chuỗi thời gian doanh thu, với quá trình chia dữ liệu như sau:

- **Tập huấn luyện:** 126 quan sát đầu tiên của chuỗi.
- **Tập kiểm định:** 19 tháng cuối cùng của chuỗi dữ liệu.

2.3.2 Kiểm định và xử lý tính dừng

Trước khi xây dựng mô hình, chuỗi thời gian được kiểm định tính dừng thông qua kiểm định Augmented Dickey-Fuller (ADF). Kết quả ban đầu cho thấy:

- $p\text{-value} = 0.903$, cho thấy chuỗi không dừng.
- Thống kê ADF = -0.440 , thấp hơn các giá trị tới hạn ở mức ý nghĩa:
 - 1%: -3.481 ,
 - 5%: -2.884 ,
 - 10%: -2.579 .
- Số độ trễ được lựa chọn: 12, với tổng số quan sát là 132.

2.3.3 Xác định tham số mô hình

Việc lựa chọn tham số cho mô hình được thực hiện qua hai bước chính:

1. Phân tích sơ bộ cấu trúc mô hình thông qua ACF và PACF với 20 độ trễ.
2. Tối ưu hoá tự động sử dụng autoarima, với các thiết lập:
 - Chu kỳ mùa vụ $m = 12$ (tháng),
 - Giới hạn tham số: $p, q \leq 3$ và $P, Q \leq 2$,
 - Tiêu chí lựa chọn mô hình: AIC.

Kết quả thu được là mô hình SARIMA(3, 1, 3)(1, 0, 0)[12] với giá trị AIC đạt 4323.673.

2.3.4 Quá trình huấn luyện

Mô hình được ước lượng với các thông số sau:

- Tham số AR phi mùa vụ: $[1.1131, -0.0667, -0.4152]$,
- Tham số MA phi mùa vụ: $[-1.3571, 0.4110, 0.1502]$,
- Tham số AR mùa vụ: 0.6436,
- Phương sai σ^2 : 8.458×10^{13} .

2.3.5 Đánh giá mô hình

Để đánh giá hiệu quả của mô hình, các chỉ số thống kê sau đã được tính toán:

- Log Likelihood: -2153.836 ,
- AIC: 4323.673,
- BIC: 4346.299,
- HQIC: 4332.865.

2.3.6 Phân tích phần dư

Mô hình được kiểm định qua các chỉ số sau nhằm đánh giá tính phù hợp của phần dư:

- **Kiểm định Ljung-Box:** $Q = 0.05$ với p-value = 0.82, cho thấy phần dư có tính độc lập (với p-value > 0.05).
- **Kiểm định Jarque-Bera:** $JB = 3.81$ với p-value = 0.15, và các hệ số:
 - Độ lệch (Skewness) = 0.17,
 - Độ nhọn (Kurtosis) = 3.79.
- **Kiểm định phương sai đồng nhất:** $H = 0.27$ với p-value < 0.001.

2.3.7 Dự báo doanh thu

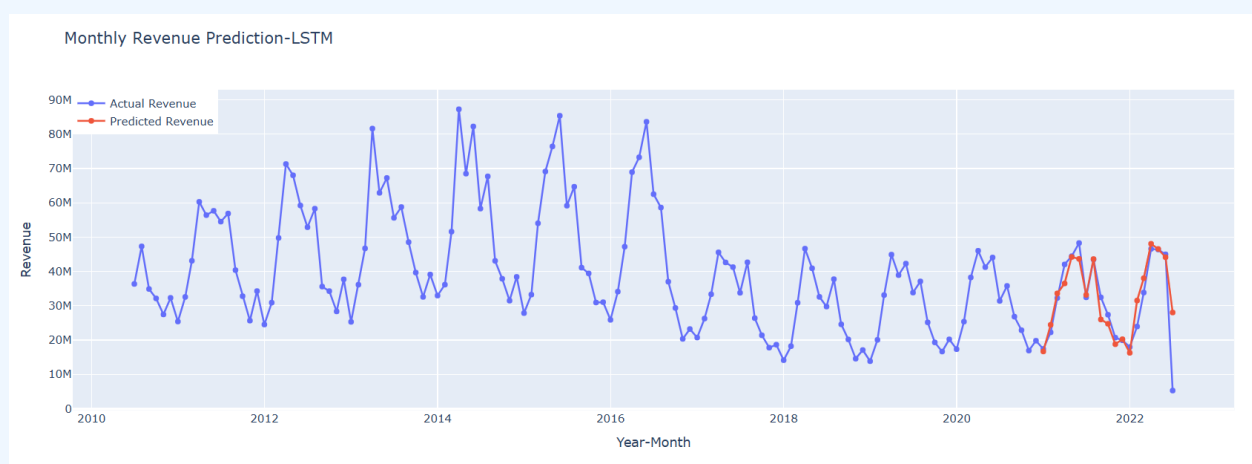
Quá trình ước lượng mô hình đạt tổng thời gian huấn luyện là 35.769 giây. Các hệ số ước lượng cho thấy:

- Hệ số AR.S.L12 có ý nghĩa thống kê rất cao (p-value < 0.001),
- Các hệ số khác có mức ý nghĩa thống kê khác nhau.

Các dự báo được thực hiện với khoảng tin cậy 95% cho từng tháng, cung cấp các giá trị dự báo doanh thu tương lai.

3 Kết quả

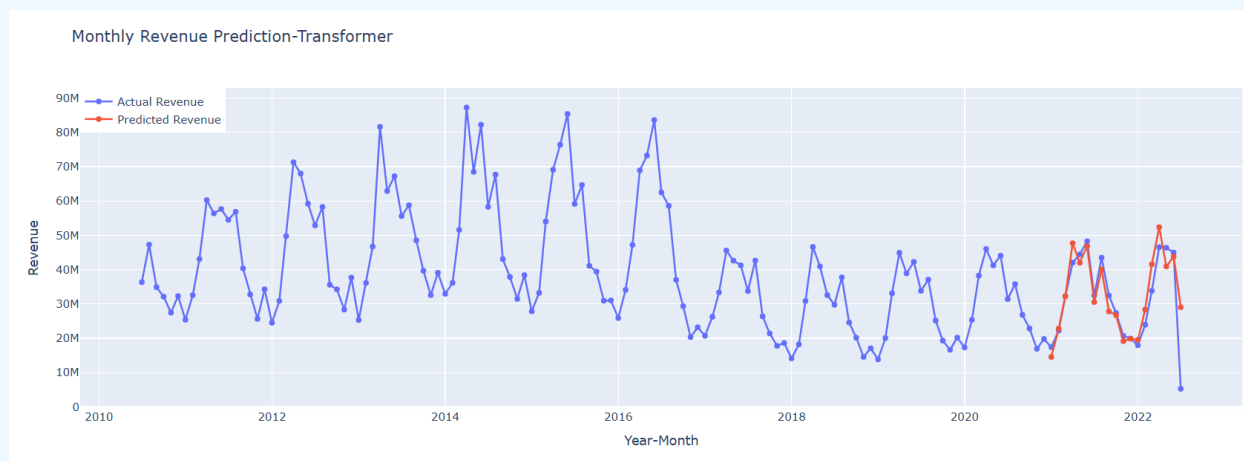
3.1 Mô hình LSTM



Hình 9: Kết quả dự đoán với mô hình LSTM

Mô hình dự đoán tương đối tốt giá trị doanh thu theo từng tháng. Tuy nhiên, có chênh lệch đáng kể tại tháng 7 năm 2022. Kết quả thực tế thấp hơn rất nhiều so với giá trị dự đoán, điều này có thể lí giải bằng việc doanh thu 2022-07 thấp đột biến so với cùng kì các năm

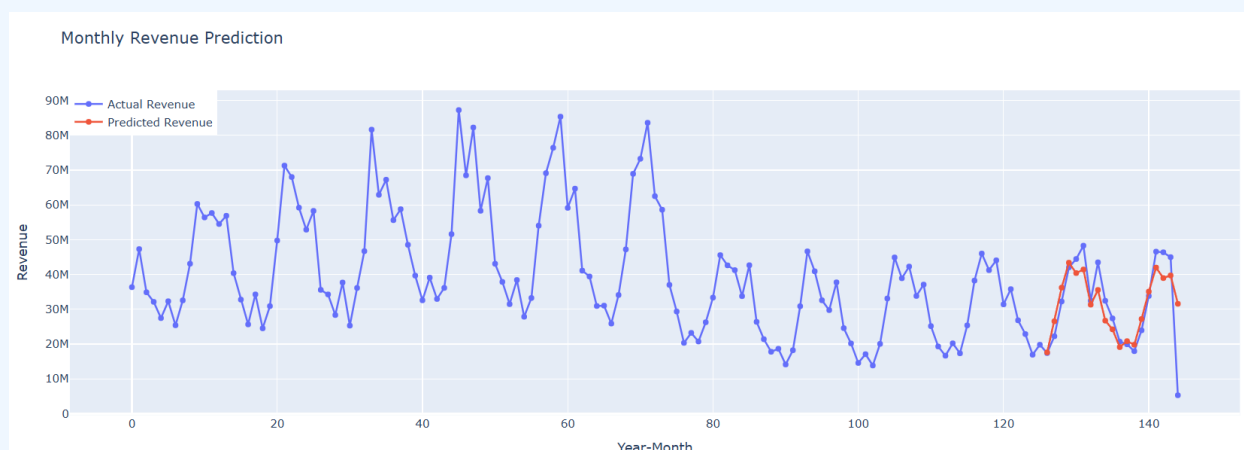
3.2 Mô hình Transformer



Hình 10: Kết quả dự đoán với mô hình Transformer

Mô hình cho kết quả dự đoán tương đối khớp với đồ thị doanh thu thực tế. Chênh lệch lớn nhất vẫn tới từ giai đoạn 2022-07, khi kết quả thực tế thể hiện mức giảm đột ngột so với cùng kì các năm và không thể dự đoán chính xác bởi mô hình

3.3 Mô hình SARIMA



Hình 11: Kết quả dự đoán với mô hình SARIMA

Mô hình SARIMA cho kết quả dự đoán tương đối khớp với dữ liệu doanh thu thực tế. Tuy nhiên, chênh lệch lớn nhất xuất hiện vào giai đoạn 2022-07, khi doanh thu thực tế giảm đột ngột so với cùng kỳ các năm trước, cho thấy hạn chế của SARIMA trong việc nắm bắt các biến động bất thường.

3.4 Đánh giá các mô hình

Mô hình	RMSE	MAE	MAPE (%)	R^2
LSTM	6,115,502.87	3,447,517.03	30.24	0.75
Transformer	6,490,637.54	3,951,829.59	31.93	0.72
SARIMA	7,328,170.25	4,789,681.21	36.10	0.64

Bảng 2: Đánh giá mô hình LSTM, SARIMA và Transformer

Cả ba mô hình đều tái hiện tương đối tốt xu hướng của dữ liệu thực tế.

LSTM đạt hiệu suất tốt nhất với sai số thấp nhất trên tất cả các chỉ số và R^2 cao nhất (0.75). Điều này cho thấy mô hình này có khả năng nắm bắt xu hướng doanh thu tốt hơn, phù hợp cho các dữ liệu có tính tuần hoàn và xu hướng dài hạn. **Transformer** hoạt động ổn định nhưng không bằng LSTM, với sai số cao hơn một chút. Mô hình này phản ứng mạnh với biến động ngắn hạn nhưng chưa tối ưu trong việc nắm bắt xu hướng dài. **SARIMA** có hiệu suất kém nhất, với sai số lớn và R^2 thấp (0.64), cho thấy mô hình này không đủ linh hoạt để theo kịp các xu hướng phức tạp trong dữ liệu doanh thu.

4 Dự đoán tình hình kinh doanh trong những năm tới

- Xu hướng chu kỳ: Dữ liệu quá khứ cho thấy doanh thu có xu hướng biến động theo mùa, với các đỉnh và đáy lặp lại khá đều đặn qua các năm. Điều này cho thấy doanh thu có tính chu kỳ.
- Dự báo tương lai: Phần dự báo tiếp tục thể hiện các biến động có chu kỳ, tuy nhiên có vẻ như mức doanh thu dự báo dao động với biên độ giảm dần so với các năm trước.
- Mức độ chính xác: Có một điểm bất thường ở cuối phần dữ liệu thực tế, khi doanh thu đột ngột giảm mạnh. Điều này có thể ảnh hưởng đến mô hình LSTM, đặc biệt nếu nó không xử lý tốt các biến động bất thường.
- Khả năng tăng trưởng: Mặc dù có biến động, nhưng phần dự báo vẫn cho thấy xu hướng ổn định, không có sự sụt giảm mạnh về doanh thu tổng thể. Điều này có thể là tín hiệu tốt nếu doanh nghiệp duy trì được các yếu tố tác động đến doanh thu. Tóm lại, trong những năm tới, doanh thu dự kiến vẫn sẽ có tính chu kỳ nhưng có thể giảm nhẹ về biên độ dao động. Cần tiếp tục theo dõi và kiểm tra mô hình để đảm bảo tính chính xác của dự báo.

Tài liệu

- [1] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.
- [2] “Predicting sales: Machine learning approaches and techniques.” <https://2001lab.io/blog/predicting-sales>. Accessed: 2025-02-24.
- [3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.