

Project: Sales Forecasting with Explainable AI (XAI)

Giảng viên: Nguyễn Thái Hà (Ph.D)

Trợ giảng: Nguyễn Thọ Anh Khoa (Ph.D Candidate)

Loại dự án: Proof of Concept (PoC)

Công nghệ sử dụng: Python | LightGBM | SHAP | Optuna | Streamlit

Source Code: https://github.com/nguyenhads/sales_forecasting_xai

Phần 1: Tổng quan dự án

Giới thiệu Tổng quan Dự án

Bài toán Kinh doanh

- **Thách thức:** Dự báo doanh số bán hàng chính xác
- **Yêu cầu:** Mô hình phải minh bạch và có thể giải thích được tại sao đưa ra dự đoán đó
- **Mục tiêu:** Hỗ trợ nhà quản lý ra quyết định kinh doanh dựa trên dữ liệu

Giải pháp Công nghệ

- **Machine Learning:** Sử dụng LightGBM - một thuật toán hiện đại để dự báo chuỗi thời gian
- **Explainable AI (AI có thể giải thích):** Sử dụng SHAP để giải thích tại sao mô hình đưa ra dự đoán cụ thể
- **Tối ưu hóa tham số:** Optuna để tự động tìm các tham số tốt nhất cho mô hình
- **Giao diện người dùng:** Streamlit tạo ứng dụng web dễ sử dụng

📄 Giải thích khái niệm:

LightGBM: Thuật toán học máy nhanh và chính xác, đặc biệt tốt với dữ liệu dạng bảng

SHAP: Công cụ giải thích mô hình AI, cho biết từng yếu tố ảnh hưởng bao nhiêu đến kết quả dự đoán

Mục tiêu dự án: **Xây dựng hệ thống dự báo doanh số bán hàng có tính minh bạch cao và dễ hiểu**



Độ chính xác (Accuracy)

Đạt được độ chính xác cao trong dự báo doanh số



Khả năng giải thích (Interpretability)

Giải thích được tại sao mô hình đưa ra dự đoán cụ thể đó



Khả năng mở rộng (Scalability)

Có thể áp dụng cho nhiều cửa hàng khác nhau



Tính khả dụng (Usability)

Giao diện dễ sử dụng cho nhân viên kinh doanh không chuyên kỹ thuật

- **Lập kế hoạch nguồn lực:** Tối ưu hóa tồn kho và nhân sự dựa trên dự báo chính xác
- **Tối ưu hóa doanh thu:** Dự báo chính xác giúp tối đa hóa doanh thu và giảm thiểu lãng phí
- **Quản lý rủi ro:** Nhận diện sớm các xu hướng bất thường trong doanh số
- **Hỗ trợ ra quyết định:** Cung cấp thông tin chi tiết cho việc lập chiến lược

Giải thích khái niệm:

- **Interpretability:** Khả năng hiểu được cách mô hình AI hoạt động và đưa ra quyết định
- **Scalability:** Khả năng hoạt động hiệu quả khi quy mô hệ thống tăng lên

Dự án tập trung vào 4 khía cạnh kỹ thuật chính: đạt hiệu suất mô hình cao, áp dụng các kỹ thuật tạo đặc trưng (feature engineering), đảm bảo khả năng giải thích rõ ràng của mô hình và sẵn sàng triển khai sản xuất.

1 Hiệu suất Mô hình

- Giảm thiểu RMSE (Sai số bình phương trung bình), MAE (Sai số tuyệt đối trung bình), MAPE (Sai số phần trăm tuyệt đối trung bình)
- Vượt trội hơn mô hình baseline Prophet

2 Kỹ thuật Tạo Đặc trưng

- Tạo ra hơn 50 đặc trưng có ý nghĩa từ dữ liệu gốc
- Đặc trưng dựa trên thời gian, lag (trễ), và thống kê của sản phẩm và cửa hàng

3 Khả năng Giải thích

- Phân tích SHAP toàn cục (ảnh hưởng chung của các yếu tố)
- Giải thích dự đoán cục bộ (giải thích từng dự đoán cụ thể)
- Xếp hạng tầm quan trọng của các đặc trưng

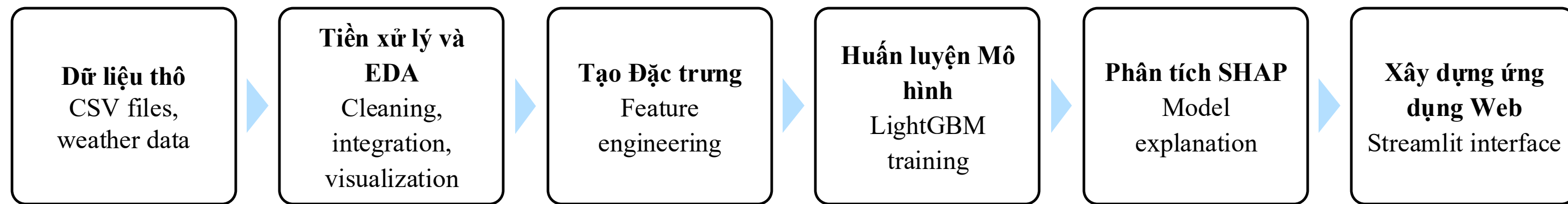
4 Sẵn sàng cho việc đưa lên production

- Code có cấu trúc module, dễ bảo trì
- Tài liệu comprehensive đầy đủ
- Triển khai thân thiện với người dùng

Giải thích khái niệm:

- **RMSE, MAE, MAPE:** Các chỉ số đo lường độ chính xác của mô hình dự báo
- **Feature Engineering:** Quá trình tạo ra các biến đầu vào có ý nghĩa từ dữ liệu thô
- **Lag Features:** Sử dụng giá trị của quá khứ để dự đoán tương lai (ví dụ: doanh số hôm qua ảnh hưởng đến doanh số hôm nay)
- **Prophet:** Sử dụng giá trị của quá khứ để dự đoán tương lai (ví dụ: doanh số hôm qua ảnh hưởng đến doanh số hôm nay)

Các bước triển khai dự án



Thư viện sử dụng

- **Xử lý dữ liệu:** Pandas (thao tác dữ liệu), NumPy (tính toán số học)
- **Học máy:** LightGBM (mô hình chính), Scikit-learn (công cụ hỗ trợ)
- **Giải thích mô hình:** SHAP (SHapley Additive exPlanations)
- **Tối ưu hóa:** Optuna (tự động tìm tham số tốt nhất)
- **Trực quan hóa:** Matplotlib, Plotly (tạo biểu đồ)
- **Giao diện web:** Streamlit (tạo ứng dụng web tương tác)

Dự án sử dụng hai nguồn dữ liệu chính là dữ liệu bán hàng và dữ liệu thời tiết, được tích hợp để phân tích và dự báo doanh số.

Dữ liệu Bán hàng

- **Phạm vi thời gian:** 2016-2017 (2 năm dữ liệu lịch sử)
- **Mức độ chi tiết:** Doanh số hàng ngày của từng cửa hàng
- **Định dạng:** Tập CSV (2016_sales.csv, 2017_sales.csv)
- **Thông tin chứa:** Ngày, mã cửa hàng, doanh số, số lượng giao dịch

Dữ liệu Thời tiết

- **Mục đích:** Phân tích tác động của thời tiết lên doanh số bán hàng
- **Thông tin:** Nhiệt độ, độ ẩm, lượng mưa, áp suất khí quyển
- **Tích hợp:** Kết hợp với dữ liệu bán hàng theo ngày và vị trí

Tích hợp Dữ liệu

- **Kết hợp datasets:** Ghép dữ liệu bán hàng + thời tiết
- **Xử lý thiếu dữ liệu:** Điền vào các giá trị bị thiếu bằng phương pháp thống kê
- **Phát hiện ngoại lệ:** Tìm và xử lý các giá trị bất thường

Cấu trúc thư mục:

```
sales_forecasting_xai/  
├── notebooks/    # 5 notebook Jupyter chính chứa toàn bộ quy trình  
├── data/         # Dữ liệu thô và đã xử lý  
├── models/       # Mô hình đã huấn luyện và các tệp liên quan  
├── src/          # Mã nguồn được tổ chức theo module  
├── figures/       # Biểu đồ SHAP và hình ảnh phân tích dữ liệu  
├── docs/         # Tài liệu hướng dẫn và báo cáo  
└── app.py        # Ứng dụng web Streamlit
```

Sản phẩm chính:

- **5 Notebook phân tích và xây dựng mô hình**
- **Mô hình LightGBM đã huấn luyện:** Sẵn sàng cho production
- **Báo cáo phân tích SHAP:** Giải thích khả năng của mô hình
- **Ứng dụng web tương tác:** Giao diện thân thiện với người dùng

Phần 2: Các bước tiến hành dự án

Bước 1 - Tiền xử lý Dữ liệu Chi tiết

Đây là bước đầu tiên và quan trọng nhất để làm sạch, chuyển đổi và chuẩn bị dữ liệu thô thành dữ liệu chất lượng cao, sẵn sàng cho việc xây dựng mô hình.

Notebook: **01_preprocessing.ipynb**



Tích hợp Dữ liệu

- Ghép nối datasets bán hàng và thời tiết
- Xử lý các tần suất dữ liệu khác nhau (hàng ngày vs hàng giờ)
- Chuẩn hóa định dạng ngày tháng



Xử lý Dữ liệu Thiếu

- Phân tích mẫu hình thiếu dữ liệu (ngẫu nhiên hay có quy luật)
- Áp dụng chiến lược điền dữ liệu phù hợp (trung bình, trung vị, interpolation)
- Kiểm tra chất lượng của việc điền dữ liệu



Phát hiện và Xử lý Ngoại lệ

- Phương pháp thống kê (IQR - Khoảng tứ phân vị, Z-score)
- Kiểm tra trực quan bằng biểu đồ
- Lọc dựa trên quy tắc kinh doanh



Đảm bảo Chất lượng Dữ liệu

- Kiểm tra tính nhất quán
- Xác thực phạm vi giá trị
- Kiểm tra logic thời gian

Giải thích khái niệm:

- **Data Integration:** Quá trình kết hợp nhiều nguồn dữ liệu khác nhau
- **Outlier:** Giá trị bất thường, khác biệt đáng kể so với phần lớn dữ liệu
- **IQR:** Thước đo độ phân tán dữ liệu, giúp phát hiện ngoại lệ

Bước 2 - Phân tích Khám phá Dữ liệu (EDA)

Đây là giai đoạn quan trọng để hiểu sâu về dữ liệu thông qua việc phân tích các mẫu hình, xu hướng và mối quan hệ ẩn chứa bên trong.

Notebook: **02_EDA.ipynb**



Mẫu hình Thời gian

- Phân tích tính mùa vụ (seasonal patterns)
- Nhận diện xu hướng dài hạn (trend identification)
- Mẫu hình chu kỳ (cyclical patterns)



Phân tích Cấp độ Cửa hàng

- Phân phối hiệu suất giữa các cửa hàng
- Gom nhóm cửa hàng theo đặc điểm tương tự
- Mẫu hình theo vùng địa lý



Tác động Thời tiết

- Phân tích mối tương quan giữa thời tiết và doanh số
- Mối quan hệ thời tiết-bán hàng
- Hiệu ứng thời tiết theo mùa



Phân phối Dữ liệu

- Tóm tắt thống kê descriptive
- Trực quan hóa phân phối
- Nhận diện bất thường



Giải thích khái niệm:

- **EDA (Exploratory Data Analysis):** Quá trình khám phá và hiểu dữ liệu trước khi xây dựng mô hình
- **Seasonality:** Mẫu hình lặp lại theo chu kỳ thời gian (hàng ngày, hàng tuần, hàng tháng)
- **Correlation:** Mức độ tương quan giữa hai biến số

Bước 3 - Kỹ thuật Feature Engineering

Đây là bước then chốt để biến đổi dữ liệu thô thành các đặc trưng có ý nghĩa, giúp nâng cao khả năng học hỏi và hiệu suất dự đoán của mô hình.

Notebook: `03_feature_engineering.ipynb`

Hơn 50 đặc trưng (features) được tạo ra

Đặc trưng Dựa trên Ngày tháng

- Thứ trong tuần, tháng, quý
- Ngày lễ, ngày cuối tuần
- Flag ngày làm việc

Đặc trưng Lag (Trễ)

- Doanh số lag 1, 7, 14, 28 ngày (doanh số từ các ngày trước đó)
- Kết hợp nhiều lag khác nhau

Thống kê cuộn (Rolling Statistics)

- Moving average cho 7, 14, 28 ngày
- Giá trị min/max, Độ lệch chuẩn
- Exponentially weighted moving average

Đặc trưng thời tiết, đặc tính cửa hàng, sản phẩm

- Nhiệt độ, độ ẩm, lượng mưa, mã hóa danh mục thời tiết
- Đặc trưng bán hàng theo cửa hàng và sản phẩm

☐ Giải thích khái niệm:

- **Feature Engineering:** Quá trình tạo ra các biến đầu vào mới từ dữ liệu gốc để cải thiện hiệu suất mô hình
- **Lag Features:** Sử dụng giá trị quá khứ để dự đoán (ví dụ: doanh số tuần trước)
- **Rolling Statistics:** Tính toán thống kê trên cửa sổ thời gian trượt (ví dụ: trung bình 7 ngày gần nhất)

Bước 4 - Phát triển Mô hình

Trong bước này, chúng ta sẽ lựa chọn và huấn luyện các mô hình dự báo, so sánh giữa mô hình baseline Prophet và mô hình nâng cao LightGBM để tìm ra mô hình tối ưu

Notebook: **04_modelling.ipynb**

Lựa chọn Mô hình:

1. Mô hình Baseline: Prophet

- Chuyên biệt cho chuỗi thời gian
- Tốt cho dữ liệu có tính mùa vụ
- Tích hợp sẵn hiệu ứng ngày lễ
- **Ưu điểm:** Dễ sử dụng, xử lý tốt missing data
- **Nhược điểm:** Ít linh hoạt với đặc trưng phức tạp

2. Mô hình Nâng cao: LightGBM

- Thuật toán gradient boosting hiện đại
- Xuất sắc với dữ liệu dạng bảng
- Huấn luyện và inference nhanh
- **Ưu điểm:** Hiệu suất cao, xử lý tốt nhiều đặc trưng
- **Nhược điểm:** Cần nhiều kỹ thuật hơn để tránh overfitting

Phân chia Dữ liệu Theo Thời gian:

- **Tập huấn luyện (Train):** Giai đoạn thời gian đầu
- **Tập xác thực (Validation):** Xác thực tuần tự theo thời gian
- **Tập kiểm tra (Test):** Giai đoạn thời gian tương lai
- **Không rò rỉ dữ liệu:** Tuân thủ nghiêm ngặt thứ tự thời gian

❏ Giải thích khái niệm:

- **Gradient Boosting:** Kỹ thuật kết hợp nhiều mô hình yếu thành một mô hình mạnh
- **Time Series Splitting:** Phân chia dữ liệu theo thời gian để tránh sử dụng thông tin tương lai để dự đoán quá khứ
- **Overfitting:** Hiện tượng mô hình học quá kỹ dữ liệu huấn luyện, kém hiệu quả trên dữ liệu mới

Bước 5 - Tối ưu hóa Hyper Parameter

Bước này giúp tự động tìm kiếm các tham số tối ưu cho mô hình LightGBM nhằm đạt hiệu suất cao nhất.

Framework: **Optuna** cho việc tìm kiếm siêu tham số (hyper parameter) hiệu quả

```
search_space = {  
    'num_leaves': (15, 50),           # Số lá trong mỗi cây  
    'learning_rate': (0.01, 0.1),     # Tốc độ học  
    'feature_fraction': (0.7, 1.0),   # Tỷ lệ đặc trưng sử dụng  
    'bagging_fraction': (0.7, 1.0),   # Tỷ lệ mẫu cho bagging  
    'bagging_freq': (1, 10),          # Tần suất bagging  
    'min_child_samples': (5, 30),     # Số mẫu tối thiểu trong lá  
    'reg_alpha': (0, 5),              # Regularization L1  
    'reg_lambda': (0, 5),             # Regularization L2  
}
```

Chiến lược Tối ưu hóa:

- **Mục tiêu**
Giảm thiểu RMSE trên tập validation
- **Cross-validation**
Cross-validation cho chuỗi thời gian
- **Số lần thử**
Hơn 100 lần thử tối ưu hóa
- **Early Stopping**
Ngăn chặn overfitting

❏ Giải thích khái niệm:

- **Hyperparameter:** Tham số không được học từ dữ liệu mà cần được thiết lập trước
- **Optuna:** Thư viện tự động tìm kiếm tham số tối ưu bằng các thuật toán thông minh
- **Regularization:** Kỹ thuật ngăn chặn overfitting bằng cách thêm penalty cho độ phức tạp của mô hình

Bước 6 - Khả năng Giải thích Mô hình

Bước này sử dụng SHAP để phân tích và giải thích cách mô hình đưa ra dự đoán, giúp tăng tính minh bạch và tin cậy của hệ thống AI.

Notebook: **05_explain_model.ipynb**



Giải thích Toàn cục (Global Explanations)

- Xếp hạng tầm quan trọng của đặc trưng
- Phân tích tác động trung bình
- Hiệu ứng tương tác giữa các đặc trưng



Giải thích Cục bộ (Local Explanations)

- Phân tích từng dự đoán riêng lẻ
- Thông tin chi tiết theo cửa hàng
- Mẫu hình theo thời gian cụ thể



Các Loại Trực quan hóa

- **Waterfall plots:** Hiển thị đóng góp từng đặc trưng vào dự đoán
- **Force plots:** Hiển thị lực đẩy/kéo của các đặc trưng
- **Summary plots:** Tổng quan tầm quan trọng các đặc trưng
- **Dependence plots:** Mối quan hệ giữa đặc trưng và dự đoán

Giải thích khái niệm:

- **SHAP (SHapley Additive exPlanations):** Phương pháp từ lý thuyết game để giải thích đóng góp của từng đặc trưng
- **Global vs Local Explanation:** Giải thích chung cho toàn bộ mô hình vs giải thích cho từng dự đoán cụ thể
- **Feature Interaction:** Cách các đặc trưng tương tác với nhau để ảnh hưởng đến kết quả

Bước 7 - Đánh giá mô hình

Bước này đo lường và xác thực hiệu suất của mô hình thông qua các chỉ số kỹ thuật và kinh doanh, đảm bảo mô hình hoạt động tốt trên dữ liệu thực tế.

Các Chỉ số Hiệu suất:

1. Chỉ số Độ chính xác:

- **RMSE** (Root Mean Square Error): Sai số bình phương trung bình
- **MAE** (Mean Absolute Error): Sai số tuyệt đối trung bình
- **MAPE** (Mean Absolute Percentage Error): Sai số phần trăm tuyệt đối trung bình

2. Chỉ số Kinh doanh:

- **Forecast bias**: Xu hướng dự báo cao hay thấp
- **Directional accuracy**: Độ chính xác về hướng thay đổi (tăng/giảm)
- **Forecast value added (FVA)**: Giá trị gia tăng so với phương pháp đơn giản

Phương pháp Xác thực:

- **Time Series Cross-Validation**: Cross-validation với cửa sổ mở rộng theo thời gian
- **Đánh giá Cấp độ Cửa hàng**: Hiệu suất riêng cho từng cửa hàng
- **Phân tích Theo Mùa**: Hiệu suất qua các mùa khác nhau
- **Phân tích Lỗi**: Nghiên cứu mẫu hình sai số còn lại

Giải thích khái niệm:

- **Cross-Validation**: Kỹ thuật đánh giá mô hình bằng cách chia dữ liệu thành nhiều phần
- **Forecast Bias**: Xu hướng hệ thống của mô hình dự báo cao hoặc thấp hơn thực tế
- **Residual Analysis**: Phân tích phần sai số để tìm hiểu điểm yếu của mô hình

Bước 8 - Phát triển Ứng dụng

Đây là bước quan trọng để chuyển đổi mô hình AI thành ứng dụng web thực tế, giúp người dùng cuối có thể dễ dàng tương tác và sử dụng hệ thống dự báo một cách trực quan.

Trực quan hóa Dữ liệu Lịch sử

- Biểu đồ chuỗi thời gian tương tác
- Lọc theo cửa hàng
- Lựa chọn nhiều khoảng thời gian

Bảng điều khiển Giải thích

- Trực quan hóa giá trị SHAP
- Phân tích đóng góp đặc trưng
- Công cụ diễn giải mô hình

Giao diện Dự đoán

- Dự báo doanh số tương lai
- Hiển thị khoảng tin cậy
- Phân tích kịch bản giả định

Trải nghiệm Người dùng

- Thiết kế giao diện trực quan
- Dự đoán thời gian thực
- Khả năng xuất kết quả

Giải thích khái niệm:

- **Streamlit:** Framework Python để tạo ứng dụng web data science một cách nhanh chóng
- **Interactive Visualization:** Biểu đồ có thể tương tác, zoom, filter
- **Real-time Prediction:** Khả năng đưa ra dự đoán ngay lập tức khi người dùng nhập dữ liệu

Tổ chức Code:

```
src/
├── data_loader/    # Module nạp dữ liệu
├── data_generator/ # Tạo dữ liệu tổng hợp
├── ui_builder/     # Thành phần giao diện
├── ui_predictor/   # Logic dự đoán
└── utils/         # Hàm tiện ích
```

Quản lý Môi trường:

- **Conda Environment:** environment.yml (cho hệ thống chung)
- **Hỗ trợ Mac M1:** environment_macm1.yml (tối ưu cho chip M1)
- **Pip Requirements:** requirements.txt (quản lý thư viện)
- **Sẵn sàng Docker:** Hỗ trợ containerization

Tính năng Mở rộng:

- **Thiết kế Module:** Dễ dàng mở rộng
- **Tham số Có thể Cấu hình:** Thiết lập linh hoạt
- **Xử lý Batch:** Dự đoán quy mô lớn

Giải thích khái niệm:

- **Containerization:** Đóng gói ứng dụng và dependencies vào container để dễ triển khai
- **Modular Design:** Thiết kế chia nhỏ thành các module độc lập, dễ bảo trì
- **Batch Processing:** Xử lý nhiều dữ liệu cùng lúc thay vì từng cái một

Kết quả: Hiệu suất Mô hình

Chỉ số Hiệu suất:

Metrics / Mô hình	① Prophet	② LightGBM	Cải thiện=① - ②	% Cải thiện
MAE	9.03	7.64	+1.39	15.39%
RMSE	11.86	11.76	+0.1	0.84%
WAPE(%)	29.13	24.82	+4.31	14.80%
Thời gian huấn luyện	yyy	yyy	+zzz	zzz%

❏ Giải thích khái niệm:

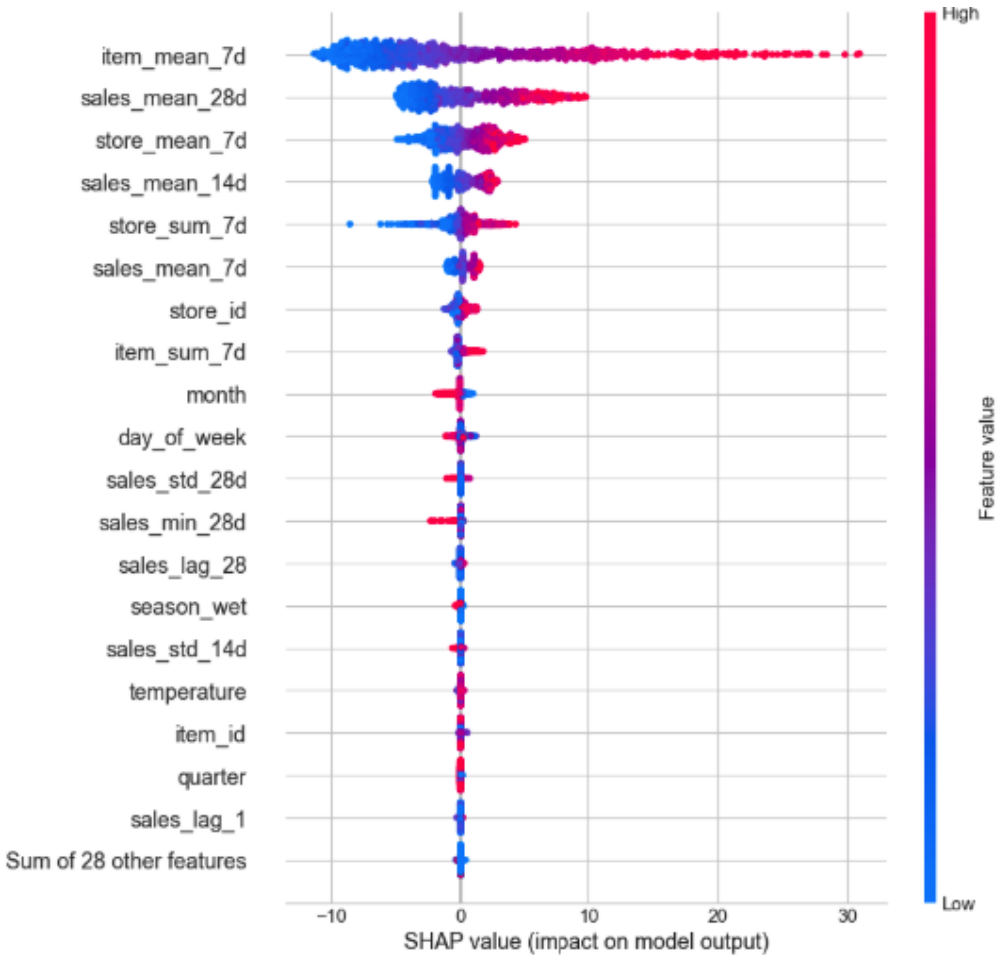
- **Baseline Model:** Mô hình tham chiếu để so sánh hiệu suất
- **RMSE, MAE, MAPE:** Các chỉ số khác nhau đo lường sai số, mỗi chỉ số có ưu điểm riêng
- **Generalization:** Khả năng mô hình hoạt động tốt trên dữ liệu chưa thấy

Kết quả: Phân tích SHAP

1. Global Feature Importance

Feature Group	Impact (%)	Explanation
Item	46.4%	Product-level features (<code>item_mean_7d</code> , <code>item_sum_7d</code> , etc.) drive forecasts.
Sales History	32.8%	Rolling features (mean, std, lag) capture recent trends effectively.
Store	18.1%	Store-level performance impacts forecasts to a moderate degree.
Date/Time	2.2%	Features like <code>month</code> , <code>day_of_week</code> have minor influence.
Weather	0.4%	Minimal effect overall.
Other	0.1%	Mostly identifiers with little or no predictive value.

Chi tiết:
https://github.com/nguyenhads/sales_forecasting_xai/blob/master/docs/shap_analysis_summary_report.md

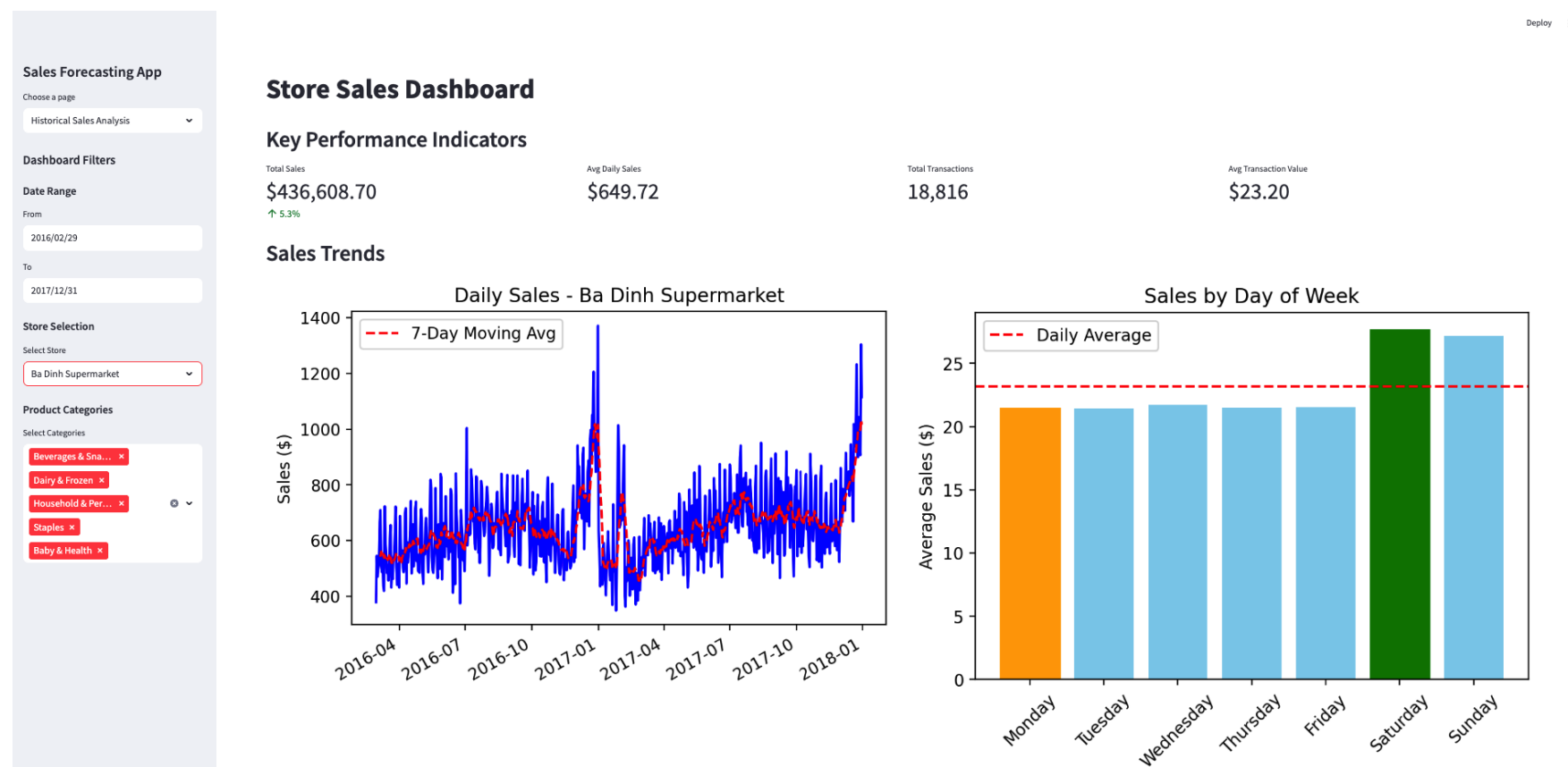


❏ Giải thích khái niệm:

- **Feature Importance:** Thứ tự ảnh hưởng của các đặc trưng đến kết quả dự đoán
- **Seasonal Pattern:** Mẫu hình lặp lại theo chu kỳ thời gian
- **Non-linear Impact:** Ảnh hưởng không tỷ lệ thuận, có thể có ngưỡng hay điểm uốn

Kết quả: Demo Ứng dụng

Giao diện ứng dụng: app.py



Trực quan hóa Dữ liệu Lịch sử

- Biểu đồ chuỗi thời gian tương tác
- Lọc theo cửa hàng
- Lựa chọn nhiều khoảng thời gian

Giao diện Dự đoán

- Dự báo doanh số tương lai
- Hiển thị khoảng tin cậy
- Phân tích kịch bản giả định

Phần 3: Tổng Kết

Thành tựu Dự án

Thành công Kỹ thuật

- **Hiệu suất mô hình:** Cải thiện so với baseline
- **AI Có thể Giải thích:** Ra quyết định minh bạch
- **Scalability:** Code có thể mở rộng và bảo trì
- **Giao diện Thân thiện:** Ứng dụng Streamlit dễ sử dụng

Tác động Kinh doanh

- **Cải thiện Dự báo:** Dự đoán doanh số chính xác hơn
- **Lập kế hoạch Nguồn lực Tốt hơn:** Tối ưu hóa quản lý tồn kho
- **Nâng cao Ra quyết định:** Thông tin dựa trên dữ liệu
- **Giảm thiểu Rủi ro:** Phát hiện sớm bất thường

Điểm nổi bật Đổi mới

- **Tích hợp XAI:** Kết hợp hiệu suất với khả năng diễn giải
- **Giải pháp End-to-End:** Pipeline ML hoàn chỉnh
- **Tech Stack Hiện đại:** Áp dụng best practices
- **Tài liệu Toàn diện:** Sẵn sàng chuyển giao kiến thức

Metrics Thành công:

ROI

Tiết kiệm chi phí

Từ dự báo chính xác



Time-to-Insight

Giảm thời gian từ dữ liệu đến quyết định



User Adoption

Mức độ sử dụng cao từ business users



Model Reliability

Hiệu suất ổn định theo thời gian

NEXT:

Cải thiện Mô hình

- **Ensemble methods:** Kết hợp nhiều mô hình (stacking, blending)
- **Deep learning:** Tích hợp LSTM, Transformer cho time series
- **Cập nhật thời gian thực:** Mô hình tự học từ dữ liệu mới

Mở rộng Feature Engineering

- **Dữ liệu ngoài:** Tích hợp chỉ số kinh tế, social media sentiment
- **Advanced time series features:** Wavelets, Fourier transforms
- **Cross-store features:** Đặc trưng từ mối quan hệ giữa các cửa hàng

Mở rộng Khả năng Giải thích

- **Counterfactual explanations:** "Điều gì sẽ xảy ra nếu..."
- **Causal inference:** Phân tích quan hệ nhân quả
- **Interactive explanation tools:** Công cụ giải thích tương tác

Mở rộng Kinh doanh

- **Scaling Multi-Store:** Triển khai doanh nghiệp lớn
- **Real-time Predictions:** Tích hợp dữ liệu streaming
- **Advanced Analytics:** Hệ thống gợi ý, optimization
- **API Development:** Kiến trúc hướng dịch vụ

📌 Giải thích khái niệm:

- **Ensemble Methods:** Kết hợp nhiều mô hình để có kết quả tốt hơn từng mô hình riêng lẻ
- **Streaming Data:** Dữ liệu được xử lý liên tục theo thời gian thực
- **Causal Inference:** Xác định mối quan hệ nhân quả thực sự, không chỉ tương quan
- **API (Application Programming Interface):** Giao diện cho phép các ứng dụng khác sử dụng chức năng của hệ thống

Cảm ơn các
đã theo dõi!



※ Hình ảnh được tạo ra do GenAI