

ỨNG DỤNG ML TRONG KINH DOANH

DỰ BÁO DOANH THU HÀNG TUẦN

Nhóm thực hiện: Nhóm 3

OUR TEAM

01.

A46219_Lê Hải Sơn

02.

A46456_Đỗ Hoàng Hiệp

03.

A42264_Phạm Duy Đạt

04.

A43088_Kim Văn Toản

NỘI DUNG CHÍNH



01

Giới thiệu



02

Dữ liệu và
tiền xử lý



03

Phương pháp
tiếp cận



04

Kết quả và
kết luận

GIỚI THIỆU

Walmart là một trong những công ty bán lẻ lớn nhất và nổi tiếng nhất thế giới, được biết đến với việc cung cấp các sản phẩm đa dạng với giá cả cạnh tranh.

Walmart Recruiting - Store Sales Forecasting là một bài toán khoa học dữ liệu kinh điển tập trung vào việc xây dựng các mô hình dự đoán để dự báo doanh số trong tương lai cho các cửa hàng Walmart



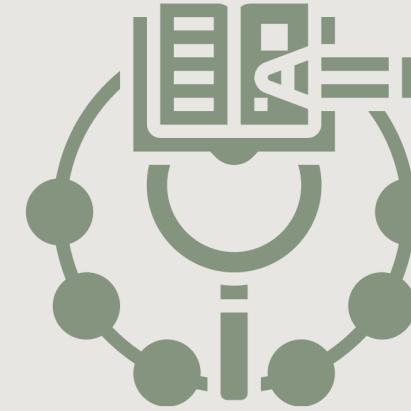
Walmart 
**Hidden
Clearance**

How To Find These
Secret Deals



Mục tiêu

- Dự đoán doanh số bán hàng tại các cửa hàng Walmart.
- Phân tích các yếu tố ảnh hưởng đến doanh số như ngày lễ, kinh tế vĩ mô và loại cửa hàng.



Kết quả

- Xây dựng mô hình dự báo doanh số chính xác với các chỉ số đánh giá hiệu suất tốt như MAE.
- Phân tích ảnh hưởng của ngày lễ và các yếu tố kinh tế đến doanh số bán hàng.



Phương pháp

- Sử dụng các kỹ thuật phân tích dữ liệu và mô hình học máy (Machine Learning).
- Áp dụng các mô hình dự báo theo thời gian (Time-series) để dự đoán doanh số.

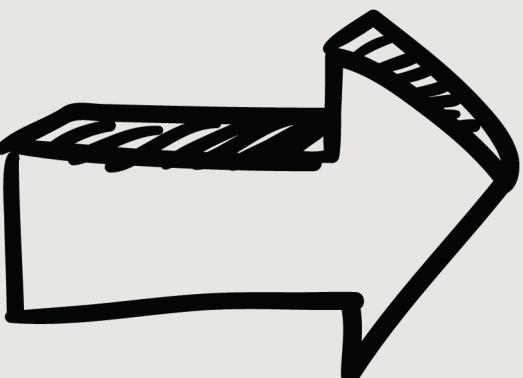
Ý nghĩa

- Giúp Walmart tối ưu hóa quản lý hàng tồn kho và nguồn lực.
- Hỗ trợ đưa ra quyết định chiến lược nhằm tăng doanh số bán hàng và nâng cao hiệu quả hoạt động.

DỮ LIỆU VÀ TIỀN XỬ LÝ

Dữ liệu gốc

Tên biến	Ý nghĩa	Kiểu dữ liệu
Store	Mã số của cửa hàng (ví dụ: Store 1)	Số nguyên
Dept	Mã số phòng ban (ví dụ: Dept 1)	Số nguyên
Date	Ngày đầu tuần (tính theo tuần)	Ngày/Thời gian
Weekly_Sales	Doanh thu hàng tuần (USD)	Số thực
IsHoliday	Có phải tuần lễ (True/False)	Boolean



Dữ liệu sau khi tiền xử lý

Tên biến	Ý nghĩa	Kiểu dữ liệu
Weekly_Sales	Doanh thu hàng tuần, biến mục tiêu để dự báo	Số thực
Date (Index)	Chỉ mục thời gian (DatetimeIndex) để xử lý chuỗi thời gian	Ngày/Thời gian

TIỀN XỬ LÝ DỮ LIỆU



Thiết lập chỉ mục thời gian

Giúp dễ dàng thao tác với dữ liệu chuỗi thời gian



Chia tập huấn luyện và kiểm tra

Đảm bảo mô hình được kiểm tra trên dữ liệu tương lai.



Kiểm tra và xử lý tính dừng

Đảm bảo dữ liệu phù hợp với yêu cầu của mô hình ARIMA.



Tạo đặc trưng thời gian

Nắm bắt các yếu tố chu kỳ và mùa vụ trong dữ liệu.

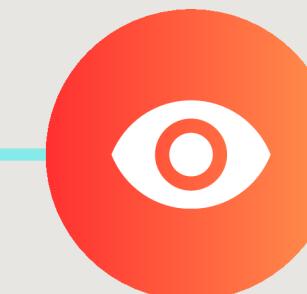
LÝ DO CHỌN ARIMA

Mô hình	Ưu điểm	Hạn chế	Lý do không chọn
ARIMA	Xử lý tốt xu hướng và tự tương quan.	Không trực tiếp xử lý mùa vụ, cần tích hợp thêm.	Được chọn vì phù hợp với dữ liệu ban đầu (chuỗi thời gian không dừng).
SARIMA	ARIMA mở rộng, xử lý tốt cả mùa vụ và xu hướng.	Phức tạp hơn ARIMA, cần xác định thêm tham số chu kỳ (seasonal order).	Chưa cần thiết nếu dữ liệu không có mùa vụ rõ ràng.

PHƯƠNG PHÁP TIẾP CẬN



Chuẩn bị dữ liệu



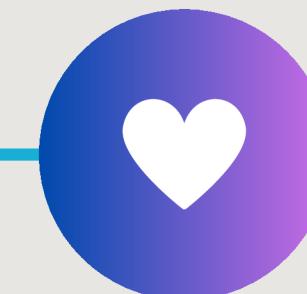
Chia tập dữ liệu

- Tập huấn luyện: **75%** dữ liệu đầu tiên.
- Tập kiểm tra: **25%** dữ liệu cuối cùng.



Huấn luyện Mô hình ARIMA

Huấn luyện trên tập huấn luyện và dự báo trên tập kiểm tra.



Đánh giá mô hình

- Sử dụng MASE (Mean Absolute Scaled Error)
- Đo lường sai số dự báo, chuẩn hóa theo độ biến động tự nhiên của chuỗi thời gian.



Trực quan hóa

- Vẽ biểu đồ so sánh giữa:
 - Dữ liệu thực tế (Actual).
 - Dự đoán từ tập huấn luyện (Predicted).
 - Dự báo trên tập kiểm tra (Forecast).

KẾT QUẢ

Hiệu quả mô hình: **MASE = 0.706**

Dự đoán chính xác trên tập huấn luyện:

Mô hình ARIMA nắm bắt tốt xu hướng trong dữ liệu

Dự báo hợp lý trên tập kiểm tra:

- Sai số dự báo (MASE) nằm trong phạm vi chấp nhận được.
- MASE < 1: Mô hình tốt hơn dự đoán naive (dự đoán giá trị hiện tại bằng giá trị trước đó).



TRỰC QUAN HÓA



1. Mô hình ARIMA hoạt động tốt trong việc dự đoán xu hướng chính:

- Đường Predicted bám sát đường Actual trong tập huấn luyện, cho thấy mô hình đã nắm bắt tốt xu hướng và dao động lớn.

2. Hạn chế trong việc dự đoán các biến động bất thường:

- Các đợt đỉnh cao hoặc thấp bất thường trong dữ liệu thực tế chưa được phản ánh chính xác trong cả tập huấn luyện và dự báo.

3. Dự báo tương lai có độ chính xác tương đối:

- Mô hình dự báo xu hướng chính xác trong tập kiểm tra nhưng thiếu chi tiết ở các biến

KẾT LUẬN

Khả năng của mô hình:

- ARIMA là công cụ mạnh mẽ để dự đoán doanh thu với các chuỗi thời gian có xu hướng và mùa vụ.
- Mô hình hoạt động tốt trên tập huấn luyện và kiểm tra, thể hiện sự tin cậy trong dự báo tương lai.

Hạn chế:

Mô hình có thể không hoạt động tốt nếu dữ liệu chứa các bất thường (spikes) hoặc các yếu tố bên ngoài (holiday effects) không được xử lý.

THANKS!