

FORECASTING BUSINESS PERFORMANCE

I. NỘI DUNG BÀI TOÁN

Dự báo kết quả kinh doanh của một công ty thời trang tại Mỹ trong 5 năm qua và dự đoán doanh thu, số lượng bán trong năm tiếp theo để hỗ trợ lập ngân sách nhập hàng.

II. MỤC TIÊU PHÂN TÍCH

Phân tích dữ liệu lịch sử, trực quan hóa xu hướng doanh thu và số lượng bán theo thời gian.

- Tạo và diễn giải các biểu đồ về xu hướng doanh số, doanh thu theo năm, tháng, quý.
- Xác định các yếu tố ảnh hưởng đến doanh thu (mùa vụ, khu vực, dòng sản phẩm).

III. YÊU CẦU

3.1. Hệ số đánh giá mô hình:

- **R-Squared (R^2):** Đánh giá độ phù hợp của mô hình
- **Mean Absolute Percentage Error (MAPE):** Đánh giá sai số trung bình phần trăm
- **Root Mean Square Error (RMSE):** Đánh giá mức độ sai số

3.2. Các yêu cầu khác

- **Báo cáo ngắn gọn (< 3 trang A4)** tập trung vào:
 - + Phân tích dữ liệu (EDA)
 - + Các tham số mô hình tối ưu đã chọn
 - + Kết quả huấn luyện và kiểm thử mô hình
 - + So sánh hiệu suất giữa các mô hình
- **Mã nguồn Python** (Jupyter Notebook / Python Script) chứa toàn bộ quá trình phân tích, xử lý dữ liệu và huấn luyện mô hình.

3.3. Mô hình dự báo

Dự đoán doanh thu và số lượng bán trong năm tiếp theo bằng các mô hình dự báo khác nhau. *Lưu ý thực hiện theo tối thiểu 2 mô hình và so sánh kết quả của 2 mô hình. Gợi ý:*

- ARIMA (Autoregressive Integrated Moving Average)
- LSTM (Long Short-Term Memory)

Ngoài ra: Nhóm có thể sử dụng mô hình khác (ví dụ: Large Language Models, Transformer, XGBoost,...) với kết quả tốt hơn ARIMA & LSTM, sẽ được cộng điểm thưởng.

Các yêu cầu cụ thể, với một mô hình, ví dụ:

- **ARIMA:**

- + Kiểm định tính dừng bằng Dickey-Fuller test
- + Sử dụng sai phân bậc 1 để làm cho dữ liệu dừng
- + Xác định giá trị tối ưu của (p, d, q) để chọn mô hình tốt nhất
- + Tiến hành dự báo

- **LSTM:**

- + Chuẩn hóa dữ liệu bằng Min-Max Scaling
- + Xây dựng mô hình LSTM
- + Huấn luyện mô hình trên dữ liệu quá khứ, dự báo doanh số và doanh thu trong năm tiếp theo
- + Tinh chỉnh hyperparameter để tối ưu kết quả

IV. MÔ TẢ DỮ LIỆU

- **Source:** Dataset nội bộ của công ty: [Forecasting Business Performance](#)

- **Period:** 10 năm gần nhất (2011 - 2020)

- **Features:**

- + **ProductID:** Mã sản phẩm
- + **Date:** Ngày giao dịch
- + **Zip:** Mã vùng bán hàng
- + **Units:** Số lượng sản phẩm bán ra
- + **Revenue:** Doanh thu

V. TIÊU CHÍ ĐÁNH GIÁ

- Độ chính xác của dự báo (các chỉ số R^2 , MAPE, RMSE)
- Chất lượng phân tích dữ liệu
- Sự sáng tạo trong cách tiếp cận mô hình