📊 Sales Forecasting Model using Supervised Learning

By

Powered by Linear Regression & Random Forest

Presented by Favour Chukwuemeka

Agenda

- Objective
- Pain Points
- Dataset Overview
- Model Approach
- Key visuals
- Feature Engineering
- Conclusion &Insights

START

Objective

This project applies supervised learning to predict sales based on product orders to support data driven decision making by using both Linear Regression and Random Forest models.

It aims to identify patterns to forecast sales effectively in an orderly behavior.

START

🔎 Summary of Business Pain Points Solved

✅ **Pain Point 1: Uncertain Sales Forecasts**

- Problem: The Businesses struggle to predict revenue.
- Solution: The model forecasts future sales by using past sales, product lines, pricing, and quantity e.t.c with 89% accuracy ($R^2 = 0.89$).

✅ **Pain Point 2: Inventory Challenges**

- Problem: Overstocking or understocking impacts profit.
- Solution: Accurate predictions using the model can help plan inventory efficiently.

START

🔎 Summary of Business Pain Points Solved

✅ **Pain Point 3: No Insight into Sales Drivers**

- Problem: It is unclear to know what affects sales performance.
- Solution: The model's feature importance shows which factors (e.g., TOTAL_VALUE, PRODUCTLINE) influence sales most.

START

📁 Dataset Overview

- ✅ Data Source: Kaggle Sales Dataset
- 📄 Features: ProductLine, QuantityOrdered, PriceEach, OrderDate, etc.
- 🎯 Target: Sales (calculated using Quantity x PriceEach)

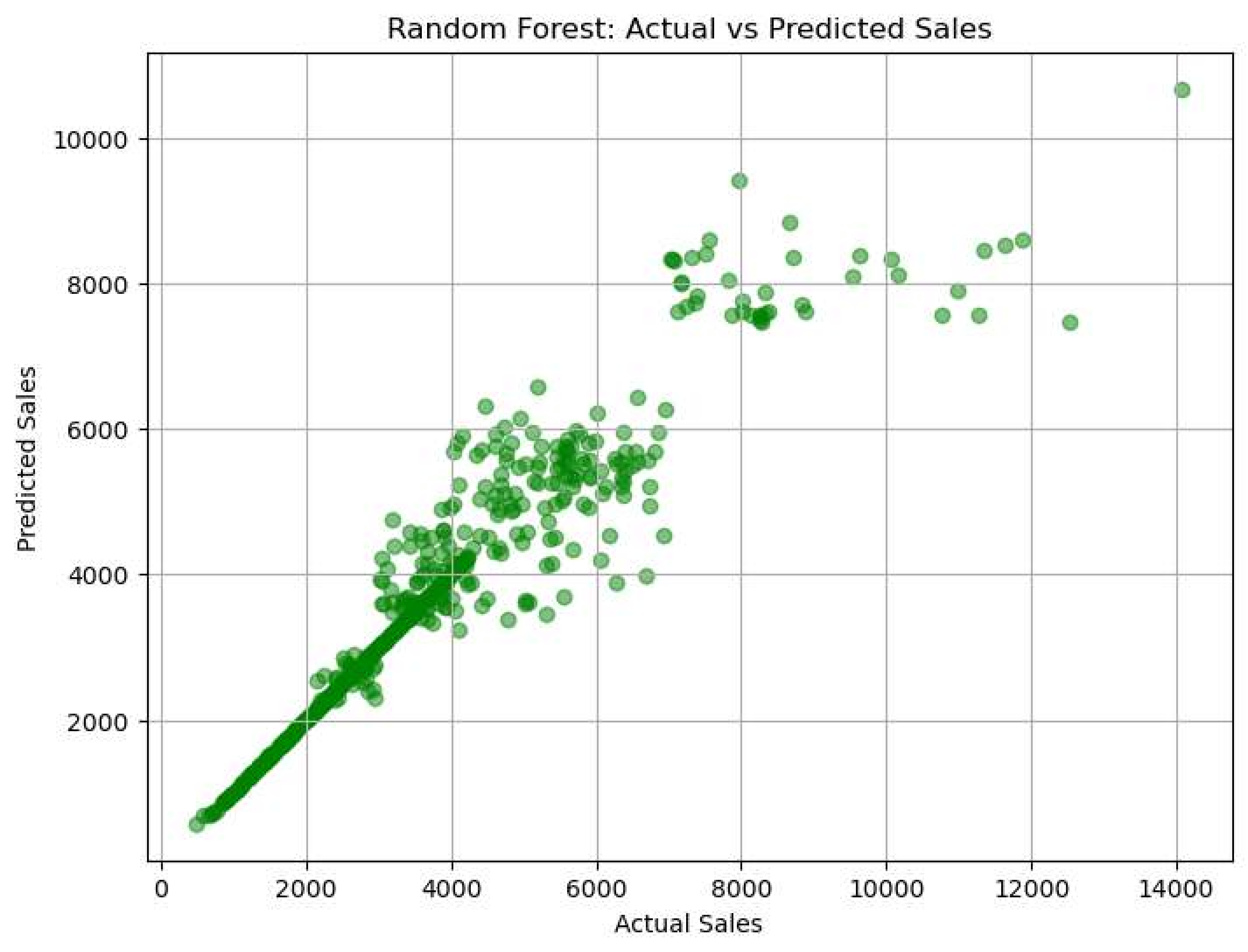
START

🧠 Model Approach

- 1️⃣ Linear Regression: Simple, interpretable baseline model.
- 2️⃣ Random Forest: More accurate, handles non-linearity better.
- 🔁 Trained/Test Split: 80/20 split used for evaluation.

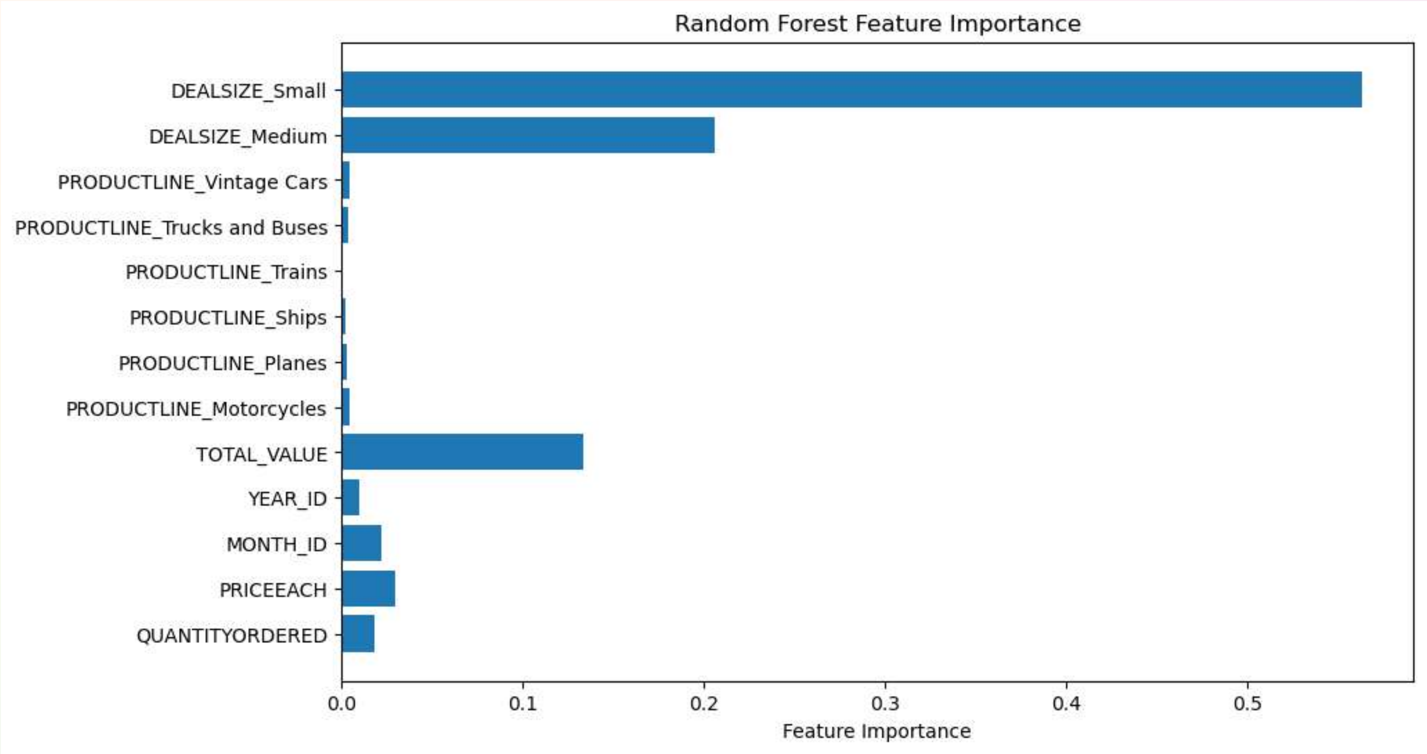
START

Key Visuals



Actual vs Predicted: Most points are close to the diagonal, meaning the model is predicting well.

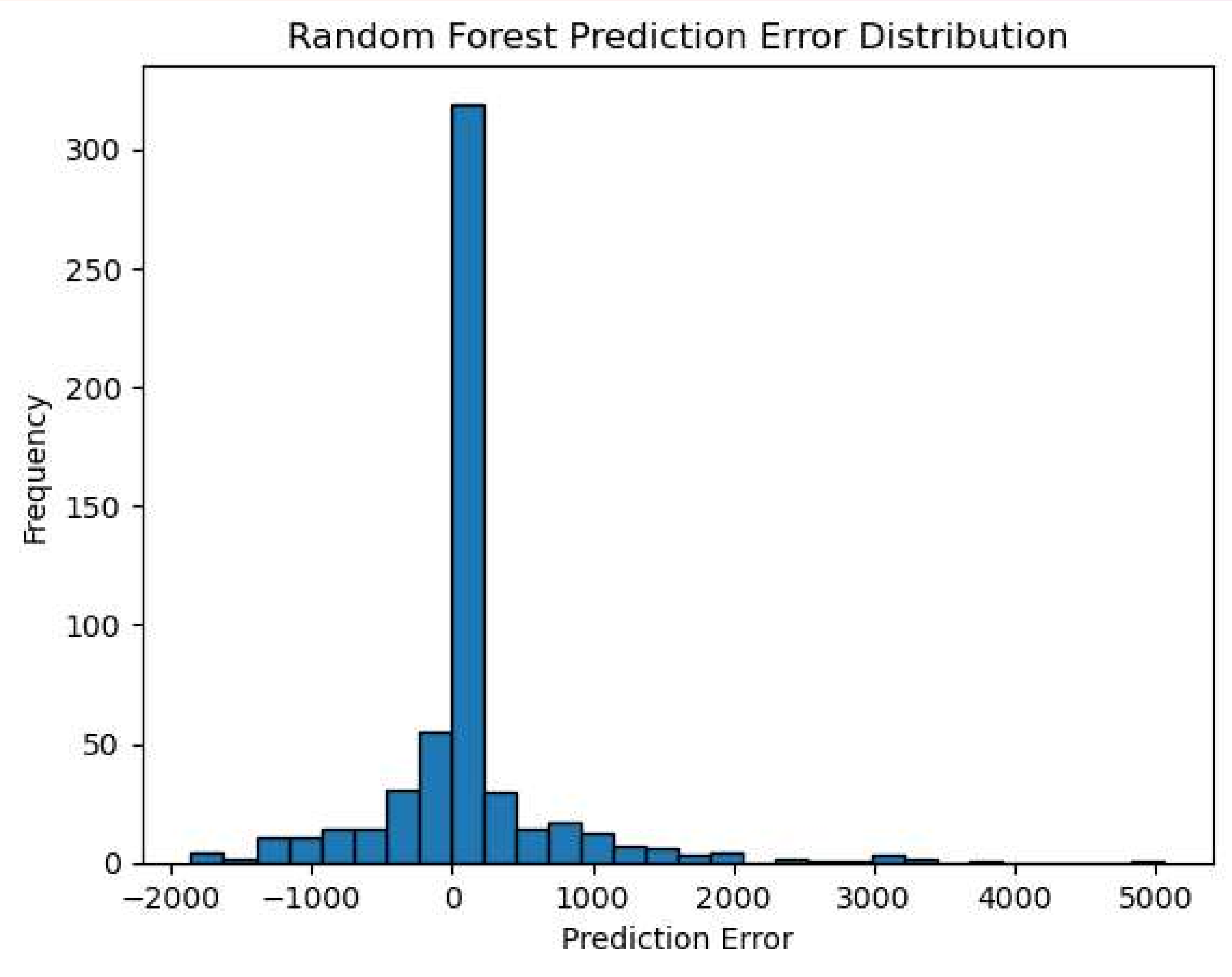
START



Feature Importance:

TOTAL_VALUE is the top feature driving prediction—logical, since it directly relates to sales.

START



Error Distribution: Most errors center around zero with a slight spread. This suggests that the model is generally accurate but still has room to improve.

START

🔧 Feature Engineering

- 🧮 TOTAL_VALUE created as QuantityOrdered × PriceEach.
- 📅 Extracted Year, Month from OrderDate.
- 🔀 Categorical variables (like ProductLine) encoded numerically.

START

📌 Conclusion & Insights

- The sales prediction model performs well, with Random Forest showing stronger results than Linear Regression.

✅ Random Forest improved prediction accuracy (MAE dropped from 469 to 326).

- 📈 R^2 of 0.89 shows strong performance on unseen data.

🔍 Business teams can rely on this for planning, stock control, and reporting

START

Thank you

Chukwuemeka Favour

START