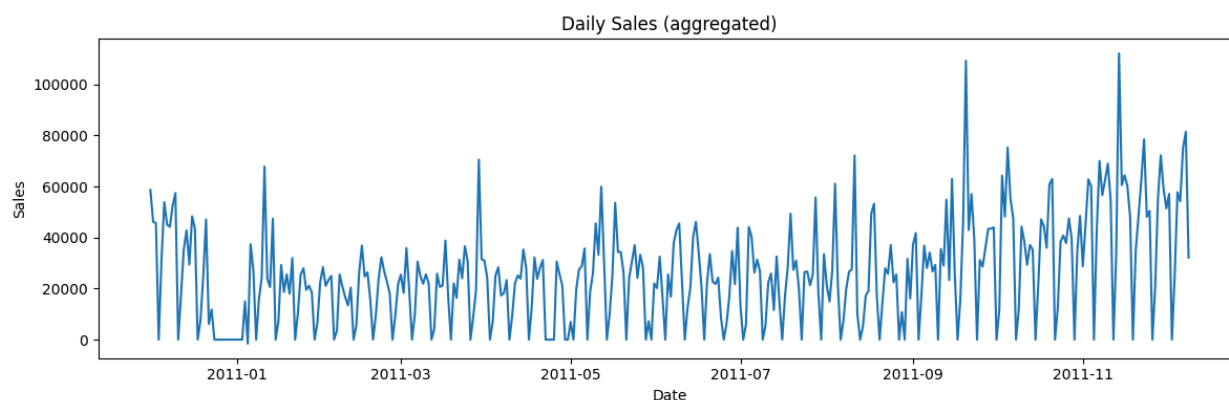


Daily Sales Forecasting Analysis – Explanation Document

This document describes a time series forecasting project on daily sales data from an online retail dataset. Two main models are used for forecasting: a classical seasonal SARIMA model and a machine learning baseline with Random Forest regression on lag features.

Daily Sales Time Series (Line Plot)

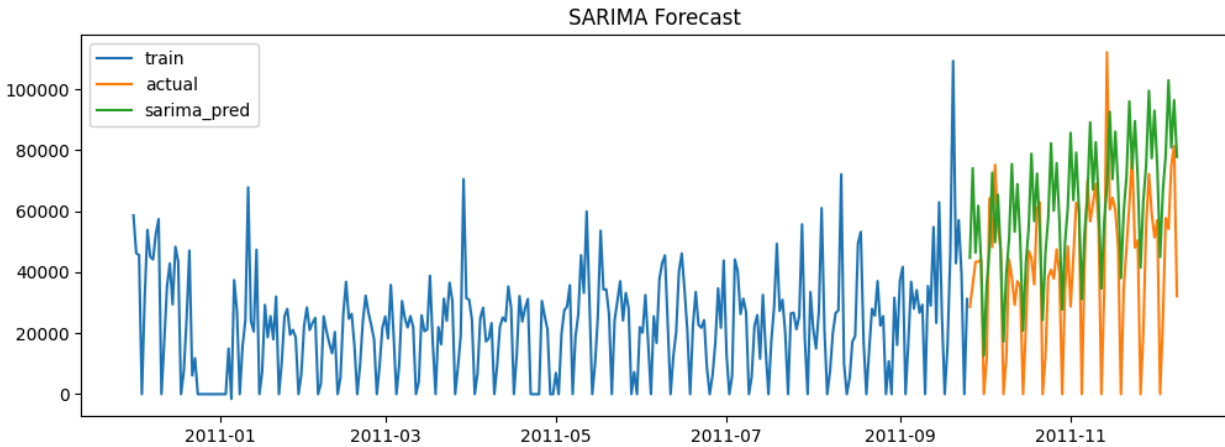


The first chart shows the daily aggregated sales over time.

- The time series reflects clear recurring patterns such as weekly seasonality and sales volatility.
- Observed variability highlights the complex nature of retail demand, affected by promotions, seasonality, and external factors.
- Visualization is essential for understanding data trends and seasonality before modeling.

Insight: Knowing the cyclicity with weekly patterns helps justify the use of a SARIMA model incorporating seasonal parameters.

SARIMA Forecast vs Actual (Line Plot)

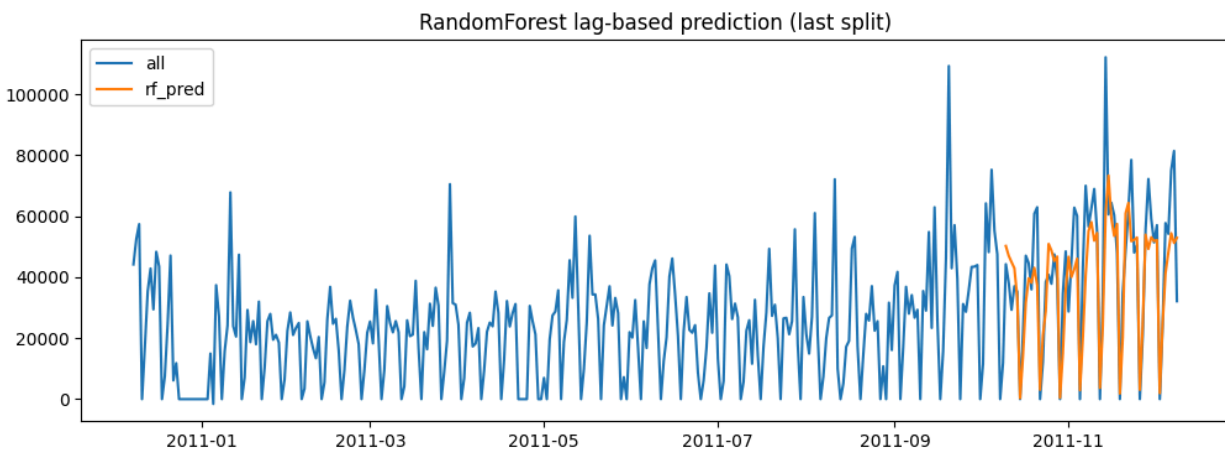


The second plot compares the SARIMA model predictions against actual sales in the testing period.

- SARIMA model was fit with parameters (1,1,1) for autoregression, differencing, and moving average components, combined with a seasonal component of weekly periodicity (7 days).
- The forecast captures main seasonal trends and fluctuations reasonably well but can smooth out sharp spikes or dips.
- SARIMA's strength is its interpretability and direct incorporation of seasonal cycles.

Insight: SARIMA adequately models weekly seasonality and trend but has limitations capturing complex nonlinearities or sudden irregular spikes.

Random Forest Prediction with Lag Features (Line Plot)



The third plot displays the Random Forest regression model's predictions based on the one-week lagged sales features compared to the actual sales.

- This machine learning approach uses previous 7 days' sales as predictors, enabling the model to discover nonlinear relationships.
- Random Forest, being a flexible ensemble method, may better capture sudden changes or complex patterns unseen by parametric SARIMA.
- The model was validated using a time series split cross-validation, ensuring future predictions depend only on past data.

Insight: The Random Forest model provides a complementary approach, especially effective when underlying dynamics are non-linear or influenced by factors not modeled in SARIMA.

Business and Modeling Implications

- Combining classic time series and machine learning models provides a robust forecasting framework, useful for inventory management and resource planning.
- SARIMA offers transparent seasonal adjustments, beneficial for understanding regular demand cycles.
- The Random Forest model adds flexibility to capture arbitrary patterns but requires careful feature engineering for lags and other variables.
- Forecast accuracy can be further improved by integrating external data and fine-tuning model parameters through grid searches or automated tools.