Sales Forecasting Project Report

By Abdul Mohiz

---

1. Introduction

Sales forecasting is a fundamental task in business analytics, helping companies predict future sales volumes for better inventory, marketing, and financial planning. This project explores time series forecasting using three powerful models:

- ARIMA (AutoRegressive Integrated Moving Average)
- SARIMAX (Seasonal ARIMA with exogenous regressors)
- Prophet (Facebook's open-source forecasting tool)

The goal is to evaluate these models on a given sales dataset and compare their performance in predicting sales amounts.

---

2. Dataset Description

The dataset contains monthly sales data with the following columns:

- SalesAmount: Monthly sales figures (target variable)
- Promotion: Binary variable indicating whether a promotion was active that month
- HolidayMonth: Binary variable indicating a holiday month

Data spans 48 months, divided into training (first 36 months) and validation (last 12 months).

---

3. Methodology

3.1 Data Preparation

- Split the dataset into training and validation sets.
- Prepare exogenous regressors (Promotion and HolidayMonth) for SARIMAX.
- Convert dates to proper time indices for Prophet.

3.2 Model Training

- **ARIMA**: Trained with parameters tuned for seasonality.

- **SARIMAX**: Included seasonal components and external regressors.

- **Prophet**: Configured to consider holidays and seasonality.

3.3 Evaluation Metrics

To evaluate forecast accuracy on validation data, the following metrics were used:

- **MAE (Mean Absolute Error)**: Average absolute difference between actual and predicted.

- **MSE (Mean Squared Error)**: Average squared difference (penalizes large errors).

- **RMSE (Root Mean Squared Error)**: Square root of MSE, in original units.

- **MAPE (Mean Absolute Percentage Error)**: Average percentage difference, useful for interpretability.

---

4. Results

Model	MAE	MSE	RMSE	MAPE (%)
ARIMA	1770.87	5,700,621	2387.60	12.96
SARIMAX	2208.45	7,139,336	2671.95	17.81

Observations:

- ARIMA provided the lowest error metrics overall, showing strong baseline performance.

- SARIMAX, despite including Promotion and HolidayMonth regressors, did not outperform ARIMA, suggesting possible regressor effectiveness or data encoding issues.

---

5. Analysis & Discussion

Model Performance

The ARIMA model, a tried-and-true classical approach, showed strong forecasting ability with an acceptable MAPE near 13%. This suggests it captures the main patterns and seasonality in sales data effectively.

SARIMAX incorporated additional domain knowledge through external regressors, but its higher error rates indicate those regressors might not have been predictive or were noisy. Further feature engineering or variable selection may improve SARIMAX.

Prophet is designed to model complex seasonality and holiday effects intuitively, but results depend on tuning and quality of holiday definitions.

Traditional Wisdom

- Simpler models often perform well if seasonality and trend are stable.

- Incorporating domain knowledge via external regressors is valuable but requires careful validation.

- Visual inspection of residuals and predictions is key to diagnosing model fit.

- Forecasting is iterative—continuous refinement based on results leads to improvement.

---

6. Conclusion

This project reinforced the strength of classical time series methods in sales forecasting. ARIMA remains a robust baseline, while SARIMAX and Prophet offer potential gains through modeling complexity and exogenous factors.

Next steps include deeper parameter tuning, refined regressors, and advanced validation techniques to further enhance forecast accuracy.