📊 Store Sales Forecasting Using Time Series Models

Submitted By:

Indra Jaiswal

Date:

February 1,2026

Table of Content:

1. Introduction

Accurate sales forecasting is a critical component of retail business operations. It enables organizations to optimize inventory management, plan promotions, allocate resources efficiently, and improve overall profitability. Poor forecasting can lead to overstocking, stockouts, and revenue lossThis project focuses on **forecasting daily store sales amounts** using **time series analysis techniques**. Multiple forecasting models were implemented and compared, including **ARIMA, SARIMAX, and Prophet**, to identify the most accurate and reliable approach. The final selected model was deployed as an **interactive Streamlit dashboard** for real-time forecasting.

---

2. Objectives

The main objectives of this project are:

1. To analyze historical daily sales data
2. To build and compare multiple time series forecasting models
3. To evaluate models using statistical and error-based metrics
4. To perform hyperparameter tuning and cross-validation
5. To deploy the best-performing model in a professional dashboard
6. To provide actionable insights for business decision-making

---

3. Dataset Description

- **Dataset Type:** Time series (daily frequency)
- **Target Variable:** sales (sales amount)
- **Date Column:** date
- **Additional Features:**
  - Promotions (onpromotion)

- Public holidays (Ecuador holidays dataset)

Sample Data

Date	Sales	On Promotion
2017-01-01	24,500	0
2017-01-02	31,200	1

---

4. Data Preprocessing

The following preprocessing steps were applied:

- Converted date column to datetime and set it as index
- Sorted data chronologically to avoid data leakage
- Checked for missing and duplicate values
- Aggregated data at the **daily level**
- Applied transformations where necessary for stationarity
- Split data into **train and test sets** using time-aware splitting

---

5. Methodology

5.1 Models Implemented

5.1.1 ARIMA (AutoRegressive Integrated Moving Average)

- Captures trend and autocorrelation
- Suitable for short-term forecasting
- Requires stationary time series

5.1.2 SARIMAX

- Extends ARIMA with:
    - Seasonal components

    - Exogenous variables (promotions)
- Weekly seasonality (7 days) was tested

5.1.3 Prophet

- Additive time series model
- Automatically models trend and seasonality
- Tested **with and without holidays**

---

6. Hyperparameter Tuning

ARIMA & SARIMAX

- Grid search over (p, d, q) parameters
- Seasonal grid (P, D, Q, s) for SARIMAX
- Model selection based on **AIC (Akaike Information Criterion)**

Prophet

- Grid search over:
    - changepoint_prior_scale
    - seasonality_prior_scale
    - holidays_prior_scale

---

7. Model Evaluation Metrics

The models were evaluated using:

- **MAE (Mean Absolute Error)**
- **RMSE (Root Mean Squared Error)**
- **AIC**
- **Time Series Cross-Validation**

---

8. Results and Model Comparison

8.1 Test Set Performance

Model	**MAE**	**RMSE**
ARIMA	**26,505.89**	**43,422.49**
SARIMAX	69,032.02	87,154.60
Prophet	37,628.94	50,568.55
Prophet + Holidays	37,628.94	50,568.55

8.2 Cross-Validation Results (MAE)

Model	**CV MAE**
ARIMA	**53,151.85**
SARIMAX	291,196.90
Prophet	75,561.84

---

9. Holiday Impact Analysis

Prophet models were trained **with and without holiday effects**.
Both models produced **identical MAE and RMSE**, indicating that:

- Holidays did **not significantly impact sales**
- The base Prophet model was sufficient
- Holiday features did not add predictive value for this dataset

---

10. Final Model Selection

✅ **Selected Model: ARIMA**

Reasons for Selection:

- Lowest MAE and RMSE
- Best cross-validation stability
- Simple and interpretable
- Efficient and production-ready

Why SARIMAX was rejected:

- Exogenous variable (promotion) did not improve accuracy
- Overfitting and unstable cross-validation results

Why Prophet was rejected:

- Higher error than ARIMA
- Holiday effects were insignificant

---

11. Deployment

The final ARIMA model was saved using joblib and deployed using **Streamlit**.

Dashboard Features:

- Interactive forecast horizon
- Daily sales amount forecasting
- Confidence intervals
- KPI metrics (latest sales, average forecast, growth)
- Top 5 highest forecasted sales days
- CSV download functionality
- Professional, business-friendly UI

---

12. Business Impact

- Improves **inventory planning**
- Helps forecast short-term demand accurately
- Reduces stockout and overstock risk
- Assists management in sales strategy planning

---

13. Limitations

- Model trained on historical patterns only
- External shocks (economic changes, pandemics) not captured
- Limited benefit from exogenous variables

---

14. Future Scope

- Weekly and monthly forecasting
- Deep learning models (LSTM, GRU)
- Automated retraining pipeline
- Multi-store forecasting
- Real-time data integration

---

15. Conclusion

This project demonstrates the effectiveness of classical time series models in retail sales forecasting. Through rigorous evaluation and cross-validation, **ARIMA emerged as the most reliable and accurate model** for short-term daily sales forecasting. The deployment of the model in a Streamlit dashboard ensures practical usability and real-world business value.
