Multivariate Time Series Forecasting of Walmart Sales Using SARIMAX

# 1. Objective

To forecast future weekly sales for a Walmart store by modeling seasonal patterns and incorporating external variables such as fuel price, temperature, unemployment rate, CPI, and holidays. The SARIMAX model is used for its ability to handle both seasonality and exogenous variables in time series forecasting.

# 2. Dataset Description

|  |  |
| --- | --- |
| Source | Description |
| Walmart Sales Data | Weekly historical sales data for various departments |
| Fuel Price | Fuel price at the time of sale |
| Holiday Indicator | Whether the week included a holiday (IsHoliday) |
| CPI | Consumer Price Index data |
| Unemployment | Weekly unemployment rates |
| Temperature | Weekly average temperature |

Time Period: 2010–2012  
Granularity: Weekly  
Target Variable: Weekly\_Sales  
External Variables: Temperature, Fuel\_Price, CPI, Unemployment, IsHoliday

# 3. Data Preprocessing

- Filtered data for Store 1, Department 1  
- Converted Date column to datetime and set it as the index  
- Sorted values chronologically  
- Ensured time index is weekly for compatibility with seasonal modeling  
- Handled any missing values using forward fill (if present)  
- Scaled IsHoliday binary feature (0/1)

# 4. Feature Engineering

- Exogenous Variables Used:  
 - Temperature  
 - Fuel\_Price  
 - CPI  
 - Unemployment  
 - IsHoliday  
- Seasonal Frequency: Weekly (seasonal period = 52)  
- No Lag Features/Moving Averages were added, as SARIMAX internally handles past dependencies

# 5. Model Selection

Models Considered:  
- Linear Regression (baseline)  
- Random Forest Regressor  
- XGBoost Regressor  
- SARIMAX (chosen) for time series forecasting with exogenous variables and seasonality  
  
Final Model: SARIMAX with configuration (1,1,1)(1,1,1,52) using exogenous variables.

# 6. Model Evaluation

|  |  |
| --- | --- |
| Metric | Value |
| MAE | ₹46,792.45 |
| RMSE | ₹73,733.17 |

SARIMAX effectively modeled the seasonal trends and responded to changes in external variables.

# 7. Visualization

1. Actual vs Forecasted Sales  
 - Clear trend alignment between real and predicted sales  
 - Plot saved as output/sarimax\_forecast\_plot.png  
  
2. Forecast Output CSV  
 - Includes date, actual sales, and forecasted sales  
 - File: output/sarimax\_forecast\_output.csv

# 8. Conclusion

- SARIMAX proved to be a robust model for multivariate time series forecasting in retail.  
- External variables such as fuel price, unemployment, and holiday flags helped capture fluctuations in demand.  
- The model can be deployed for sales planning, inventory control, and promotional strategy decisions.  
- Future work can involve testing LSTM or Prophet for comparison.