

Project Report: Walmart Store Sales Forecasting

1. Problem Statement

A retail store with multiple outlets across the country is facing challenges in managing inventory effectively. The objective is to analyze historical sales data, derive meaningful insights, and build predictive models to forecast future sales. This will help the stores optimize inventory and improve supply chain efficiency.

2. Project Objective

- Analyze historical sales data to identify trends and patterns.
- Develop predictive models to forecast sales for the next 12 weeks.
- Compare ARIMA and SARIMAX models for accuracy.
- Provide insights that can assist in better inventory management.

3. Data Description

- **Dataset Used:** Walmart Dataset
- **Key Columns:** Date, Weekly_Sales, Store, Dept, Temperature, Fuel_Price, CPI, Unemployment, IsHoliday.
- **Time Frame:** Covers weekly sales data over multiple years.

4. Data Pre-processing Steps and Inspiration

- **Missing Values Handling:** Checked and handled missing values if present.
- **Date Formatting:** Converted 'Date' column to datetime format.
- **Feature Engineering:** Created additional time-based features if needed.
- **Train-Test Split:** Data was divided into training and testing sets.

5. Choosing the Algorithm for the Project

We experimented with ARIMA and SARIMAX models:

- **ARIMA (AutoRegressive Integrated Moving Average)** for univariate time series forecasting.
- **SARIMAX (Seasonal AutoRegressive Integrated Moving Average with Exogenous Regressors)** for capturing seasonality and external factors.

6. Motivation and Reasons for Choosing the Algorithm

- ARIMA was chosen due to its efficiency in handling univariate time series.
- SARIMAX was used to incorporate seasonality and external regressors like fuel price, CPI, and unemployment rate.
- Comparing both models allowed us to choose the best-performing approach.

7. Assumptions

- Sales data follows a seasonal pattern.
- No external factors like sudden economic changes or policy changes affecting the store sales drastically.
- Past sales trends are a reliable indicator of future sales.

8. Model Evaluation and Techniques

- **Hyperparameter Tuning:** Used `itertools` method on `p`, `d`, `q` (ARIMA) and `P`, `D`, `Q`, `S` (SARIMAX).
- **Evaluation Metric:** RMSE (Root Mean Squared Error) to compare models.
- **Train-Test Split:** Last few months of data were used for testing.
- **Results Comparison:**
 - ARIMA Predictions: Produced a baseline forecast.
 - SARIMAX Predictions: Captured seasonality and performed better.

9. Inferences from the Same

- SARIMAX outperformed ARIMA due to seasonality consideration.
- Major sales spikes were observed around specific months (possibly holiday seasons).
- External factors like fuel prices and unemployment showed some correlation with sales trends.

10. Future Possibilities of the Project

- **Fine-tuning models with more external regressors.**
- **Exploring deep learning-based forecasting methods (LSTMs, Facebook Prophet).**
- **Developing a real-time sales prediction dashboard using Power BI.**
- **Applying clustering to group stores based on similar sales behavior.**

11. Model Saving

- Both ARIMA and SARIMAX models were trained and saved using Python's `pickle` or `joblib` for future use.

This report provides a structured approach to sales forecasting for the retail store. Let me know if you need any refinements!