Digital Egypt Pioneers Initiative (DEPI) Final Project

Forecasting Model Performance :Milestone (3)

Sales Forecasting and Optimization

Group Members:

Basmala Ehab Mohamed Yousry

Mostafa Mahmoud Mohamed Elshahat

Mohab Mohamed Ibrahim Mohamed

Ziad Ahmed Gharieb

SHR2_AIS4_S2

Submission Date:
9 May 2025

Table of Contents

Model Selection 1

Model Training 1

Model Evaluation and Tuning 2

Forecasting Model Performance Report 2

Final Forecasting Model: 2

Model Selection

For our sales forecasting task, we explored both statistical and machine learning models to capture various temporal and nonlinear patterns in our data. The models selected were:

- **ARIMA (AutoRegressive Integrated Moving Average)**: To capture linear trends and seasonality in time series.
- **SARIMA (Seasonal ARIMA)**: Extended ARIMA with seasonal components to handle periodic patterns.
- **XGBoost (Extreme Gradient Boosting)**: A powerful ensemble model used to capture nonlinear interactions and complex patterns in sales data.
- **Prophet (by Facebook)**: A time series forecasting model suitable for daily data with multiple seasonalities and holiday effects.
- **Linear Regression**: As a baseline machine learning approach to model the relationship between time and sales.
- **Logistic Regression**: Though typically used for classification, it was tested for binary sales outcomes (e.g., high vs. low demand) as an exploratory approach.

Model Training

- Each model was trained using historical sales data, which was split into training and validation sets. The time-series models (ARIMA, SARIMA, Prophet) used date-based indexing, while machine learning models (XGBoost, Linear, and Logistic Regression) were trained using engineered features including lagged sales, day-of-week indicators, and promotional flags.

Model Evaluation and Tuning

Models were evaluated based on standard forecasting metrics including:

- **MAE (Mean Absolute Error)**
- **RMSE (Root Mean Squared Error)**

Hyperparameter tuning was performed as follows:

- **ARIMA/SARIMA**: Parameters (p, d, q) and (P, D, Q, s) were selected using AIC/BIC scores and grid search.
- **XGBoost**: Grid search and cross-validation were applied to tune max_depth, learning_rate, and n_estimators.
- **Prophet**: Holidays, changepoints, and seasonalities were adjusted for optimal performance.
- **Linear/Logistic Regression**: Regularization techniques (L1/L2) were tested for generalization improvement.

Forecasting Model Performance Report

The performance of the models was summarized as follows:

Model	MAE	RMSE	Comments
ARIMA	677.87	88935.096	Stable on linear trends, underperforms on seasonality
SARIMA	5575.2	9776.7	Best at modeling seasonality and trend
XGBoost	137	897.86	Outperformed traditional models with engineered features
Prophet	244.22	26353.7	Easy to tune, performs well on multiple seasonal patterns
Linear Regression	50.36358	18847.08	Simple, acts as baseline model

Final Forecasting Model

After comparison, **[state final selected model, XGBoost]** was selected as the final model due to its superior performance in capturing [linear/seasonal/nonlinear] patterns with the lowest forecast error. The final model was used to generate forecasts for the next period and will be deployed for production monitoring and periodic retraining.