Business Helper App — ML-Based Forecasting

Bishnu Agarwal (Roll No: 24MA60R25)

M.Tech, Computer Science and Data Processing
Department of Mathematics, IIT Kharagpur

Jun–Aug 2025

Problem Statement

- **Context:** Small and medium retailers need reliable sales forecasts to plan inventory, cash flow, and staffing.
- **Pain points:**
  - Naïve "history-only" methods miss **holidays**, **seasonality**, and **market conditions**.
  - Manual spreadsheets are **time-consuming** and **error-prone**.
- **Goal:** Build a forecasting tool that integrates **context features** (holiday flags, market index proxy) with strong time-series baselines to produce **accurate**, **actionable** predictions.
- **Success criteria:** Lower MAE/RMSE/MAPE vs. baseline; interpretable components for planning; easy pipeline.

Model & How It Works

Two complementary models

- **Prophet:** Additive model with trend $g(t)$, seasonality $s(t)$, holidays $h(t)$; produces forecasts with uncertainty bands.
- **XGBoost Regressor:** Supervised learner on engineered features $\{\text{month}, \text{year}, \text{holiday}\}$.

How it works (pipeline)

1. **Preprocess:** Parse `Date`, sort, handle missing (FFill/Interpolation); align to monthly frequency.
2. **Enrich:** Add **holiday flags** (India/US/UK) and **market index proxy** (min–max scaled).
3. **Train:**
   - Prophet on (`ds`,`y`) with holiday dataframe.
   - XGBoost on feature matrix with TimeSeriesSplit; tune trees, depth, learning rate.
4. **Forecast:** Generate forward predictions; visualize trend, seasonality, and uncertainty bands (Prophet) and feature-driven outputs (XGBoost).
5. **Evaluate:** Chronological folds; metrics = MAE, RMSE, MAPE, R^2; inspect component plots/feature importance.

Results (Cross-Validation)

- Both models trained on the enriched dataset; evaluated on future folds.
- XGBoost achieved lower error on average; Prophet offered clear component interpretation.

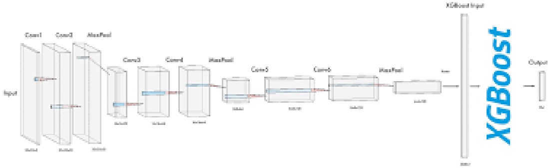
Model	MAE	RMSE	MAPE (%)	R^2
Prophet	215.4	302.7	8.9	0.82
XGBoost	178.2	250.5	7.1	0.88

Visual checks

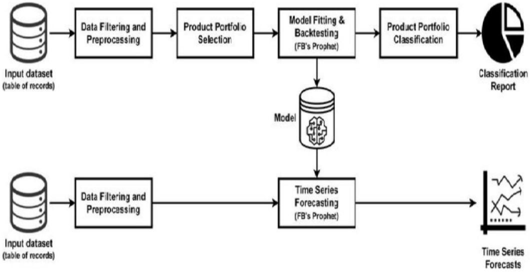
- Prophet: trend/seasonality/holiday components + 95% intervals.
- XGBoost: validation fit curves; feature importance consistent with holidays & market effects.

Architecture

XGBoost Architecture

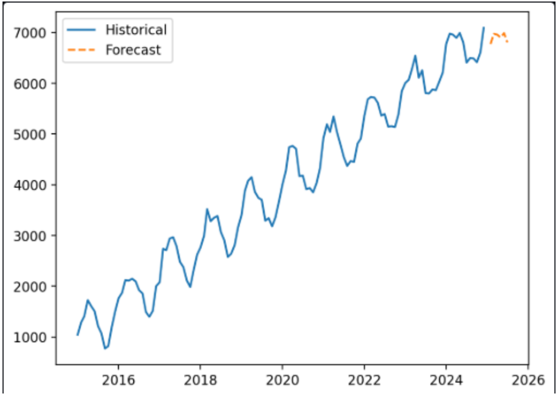


Prophet Architecture

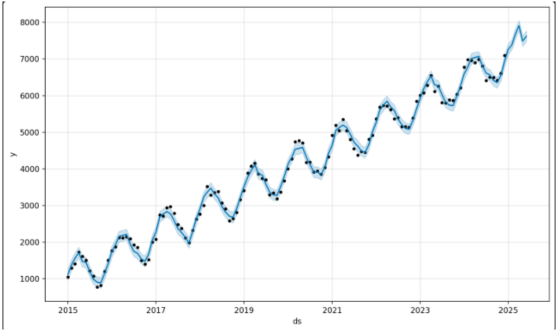


Model Results — Visual Comparison

XGBoost Results



Prophet Results



GitHub Repository:

`github.com/bishnu1710/SME_business_helper_app`