Advanced Time Series Forecasting for Demand Prediction

---

1. Title

Prepared by:

- **Suraj Prasanna** – 225890296 (AI)
  suraj.mitblr2022@learner.manipal.edu

Institution/Organization:
Department of Computer Science and Engineering
Manipal Academy of Higher Education

Bengaluru, India

---

2. Abstract

Accurate demand forecasting is crucial for optimizing supply chains and making data-driven business decisions. This project develops an industry-ready time series forecasting model using real-world sales data from a globally recognized technology company. Leveraging advanced machine learning techniques, I implemented a hyperparameter-optimized Prophet model, an XGBoost-based approach, and a hybrid ensemble method to enhance predictive accuracy.

A rigorous automated hyperparameter tuning process was employed to identify the optimal configuration for Prophet, ensuring minimal forecasting error. Additionally, statistical decomposition techniques were applied to analyze historical trends and seasonality, reinforcing model interpretability. The final forecasting framework provides robust and scalable demand predictions, offering valuable insights for inventory management, production planning, and strategic decision-making.

---

3. Introduction

Importance of Demand Forecasting

In today's competitive market, demand forecasting plays a crucial role in supply chain management, inventory optimization, and strategic planning. Businesses rely on accurate sales predictions to align production, procurement, and distribution strategies, ensuring seamless operations and cost efficiency.

Key benefits of effective demand forecasting:

- Optimized Inventory Management – Reduces overstocking and stockouts.
- Better Resource Allocation – Helps businesses allocate manpower and logistics efficiently.
- Improved Financial Planning – Enhances budgeting, pricing strategies, and revenue predictions.

Consequences of Inaccurate Predictions

Poor demand forecasting can have severe operational and financial repercussions:

- Stock shortages – Leads to lost sales and customer dissatisfaction.
- Excess inventory – Increases storage costs and capital blockage.
- Supply chain disruptions – Creates inefficiencies in procurement and logistics.
- Revenue loss – Misalignment between production and demand results in financial instability.

These challenges highlight the need for a robust, data-driven forecasting approach that minimizes errors and enhances decision-making.

The Forecasting Challenge

Forecasting demand for multiple products with limited historical data presents a unique challenge. Traditional time series models often struggle to:

- Capture seasonality and long-term trends effectively.
- Adapt to market fluctuations and product life cycles.
- Provide reliable predictions when data is scarce or irregular.

Methodology / Approach

This section details the structured methodology followed in developing a robust, industry-relevant demand forecasting model, integrating best practices in data preprocessing, model selection, and hyperparameter tuning.

Comprehensive Data Preprocessing

Effective data preprocessing is the foundation of any reliable forecasting model. Several essential steps were taken to prepare the dataset for high-quality predictions, following best practices used by professional data analysts:

1. **Data Cleaning & Feature Selection** – The raw dataset contained unnecessary columns and redundant information. A systematic filtering process was applied to retain only product names, timestamps, and sales data, ensuring a streamlined dataset.

2. **Handling Missing Values** – Initial missing values were imputed using proportional estimation based on early observed values, preventing skewed forecasting trends. Alternative imputation strategies (such as mean/moving average imputation) were considered but rejected in favor of domain-specific logic.
3. **Data Normalization & Transformation** – While the dataset did not require extensive scaling, log transformation and differencing techniques were explored to check for stationarity improvements. These are standard transformations used in professional forecasting workflows.
4. **Time Series Structuring & Indexing** – Proper datetime indexing was applied to preserve temporal consistency, ensuring accurate trend analysis. Time-based feature engineering (lagged features, moving averages, rolling statistics) was considered but excluded for interpretability purposes.

2. Model Selection & Forecasting Techniques

To ensure robust and reliable demand predictions and also because of the dataset involving products of varying trend and seasonality , and variety of product life cycles, a multi-model approach was adopted, leveraging a combination of statistical modelling and machine learning. For each time series, the model with the highest accuracy is selected.

2.1 Prophet – Capturing Trend & Seasonality

Prophet was chosen due to its ability to automatically detect long-term trends, seasonality, and holiday effects. Professional analysts Favor Prophet for:

1. Handling irregular sales fluctuations without requiring complex feature engineering.
2. Accounting for seasonal demand cycles inherent to quarterly sales.
3. Providing interpretable components for business decision making (trend, seasonality, residual analysis).

A structured hyperparameter tuning strategy was implemented to optimize:

1. **Changepoint flexibility** – Preventing overfitting by controlling how frequently trends shift.
2. **Seasonality components** – Fine-tuning yearly and quarterly seasonality to maximize accuracy.
3. **Growth model type (linear/logistic)** – Selecting the best trend modelling approach based on product behaviour.
4. **Holiday effects** – Incorporating holiday influences to capture demand fluctuations during special events and peak periods.
5. **Cross-validation optimization** – Applied a structured hyperparameter tuning strategy to enhance model performance:
    1. Iterated over a range of changepoint_prior_scale and seasonality_prior_scale values.
    2. Utilized a cross-validation scheme with an initial training window of 730 days**,** followed by a 91-day period and a 91-day forecast horizon.
    3. Selected optimal hyperparameters based on the lowest Mean Absolute Percentage Error (MAPE)**.**

2.2 XGBoost – Machine Learning for Time Series with Optuna Optimization

To complement Prophet's statistical approach, XGBoost, a gradient boosted decision tree model, was employed to capture non-linear patterns and complex relationships within the sales data. XGBoost is widely used in demand forecasting due to its:

1. Ability to detect sudden changes in demand.
2. Feature importance evaluation, aiding business insights.
3. Robustness against missing data and noise.

To maximize XGBoost's predictive performance, Optuna, an advanced hyperparameter optimization framework, was leveraged. Optuna's automated search was used to fine-tune:

1. **Learning rate** – To balance speed and accuracy.
2. **Number of estimators** – To prevent overfitting while maintaining predictive power.
3. **Max depth of trees** – To control model complexity.
4. **Regularization parameters** – To improve generalization across different products.

Optuna's ability to efficiently explore the hyperparameter space enabled the selection of the most optimal XGBoost configuration, significantly improving forecast accuracy.

2.3 Hybrid Model – Merging Statistical & Machine Learning Approaches

Recognizing that no single forecasting model is universally optimal, a hybrid strategy was implemented by combining Prophet and XGBoost predictions. This approach ensured:

1. Statistical rigor from Prophet's trend analysis.
2. Machine learning-driven adaptability from XGBoost.
3. Reduced variance by leveraging multiple predictive perspectives.

3. Model Optimization & Forecast Refinement

To ensure business-ready forecasting, rigorous validation and optimization techniques were applied:

1. **Cross-Validation Strategies** – While time-series k-fold cross-validation was considered, it was deemed impractical due to data constraints. Instead, a **sliding window validation** approach was implemented to assess forecasting performance dynamically.
2. **Error Metric Selection** – The Mean Absolute Error (MAE) was chosen as the primary evaluation metric, as it provides an intuitive and business-relevant measure of forecasting accuracy, minimizing the impact of extreme values.
3. **Automated Hyperparameter Tuning** –
    - A custom grid search function iterated through multiple Prophet configurations to identify optimal changepoint flexibility and seasonality settings.

- **Optuna optimization** was employed for XGBoost, efficiently tuning parameters such as learning rate, tree depth, and boosting parameters to align with product-specific demand patterns.

4. **Forecast Visualization & Interpretation** – Forecasted trends were integrated with historical sales data, allowing for a clear comparative analysis of past trends versus projected demand. This step ensured that forecasts were not only accurate but also interpretable and actionable for business stakeholders.

Results and Analysis

1. Overview

This section presents a detailed evaluation of the forecasting models - Prophet, XGBoost, and Hybrid based on their predictive accuracy. We utilize Mean Absolute Percentage Error (MAPE) as the primary performance metric across 20 different products. Additionally, graphical visualizations illustrate the models' forecasting trends, and a comparative bar chart provides insights into the alignment between actual and predicted sales.

2. MAPE Evaluation

Table 1 provides a comparative analysis of the MAPE values for each product across all models. The best-performing model for each product is highlighted, showcasing the lowest MAPE value, which indicates the highest predictive accuracy.

Key Findings from MAPE Analysis:

- **XGBoost** outperforms the other models in case of **products with sustaining** life cycle.
- **Hybrid** models provide better accuracy in scenarios where individual models show higher errors.
- **Prophet** performs well for **products with declining and NPI life cycles** respectively categories but exhibits inconsistencies in others.
- **Lower MAPE values indicate reliable forecasts**
  In cases where the MAPE values are significantly lower, the respective models have generated highly reliable predictions, making them suitable for operational decision-making. Conversely, higher MAPE values suggest the need for further fine-tuning or alternative forecasting strategies.
- **The need for dynamic model selection**
  Given the variance in model effectiveness across different product life cycles, a dynamic forecasting approach should be considered, where models are chosen adaptively based on the nature of the product and historical data trends.
- **Insights for business decision-making**
  The findings from this analysis can help in inventory management, resource allocation, and demand planning by identifying which models provide the most accurate forecasts for different product types.

Table 1: MAPE Evaluation of Models Across Products

Product	Actual	P (%)	X (%)	H (%)	Best
SWITCH High1	25000	24.7	**0.26**	9.7	X
SWITCH Ultra High1	6900	81.1	**0.56**	32.7	X
SWITCH Ultra High2	7200	34.5	22.5	**0.31**	H
TRANSCEIVER Mid1	62000	36.0	23.2	**0.48**	H
POWER High	84000	**0.05**	39.6	23.7	P
TRANSCEIVER High1	10900	62.1	**0.24**	24.9	X
ACCESS Mid	75000	**0.69**	32.1	19.0	P
POWER Mid	105000	5.7	4.1	**0.18**	H
PROCESSOR	1250	47.6	**0.28**	19.2	X
TRANSCEIVER Mid2	20000	**0.51**	21.6	12.8	P
SWITCH Mid2	2300	24.7	19.2	**1.65**	H
SWITCH Low	15000	**0.00**	46.6	23.3	P
SERVER	7000	14.2	21.4	**0.00**	H
SWITCH High2	2000	40.0	**0.00**	20.0	X
ROUTER Mid1	13000	23.0	23.0	**0.00**	H
MEMORY	18000	**0.00**	33.3	16.6	P
ROUTER Mid2	6000	16.6	25.0	**0.00**	H
SWITCH Mid1	1200	50.0	**0.00**	25.0	X
ROUTER Low	2500	**0.00**	60.0	28.0	P
SWITCH High3	200	100.0	**0.00**	50.0	X

Table with MAPE Values

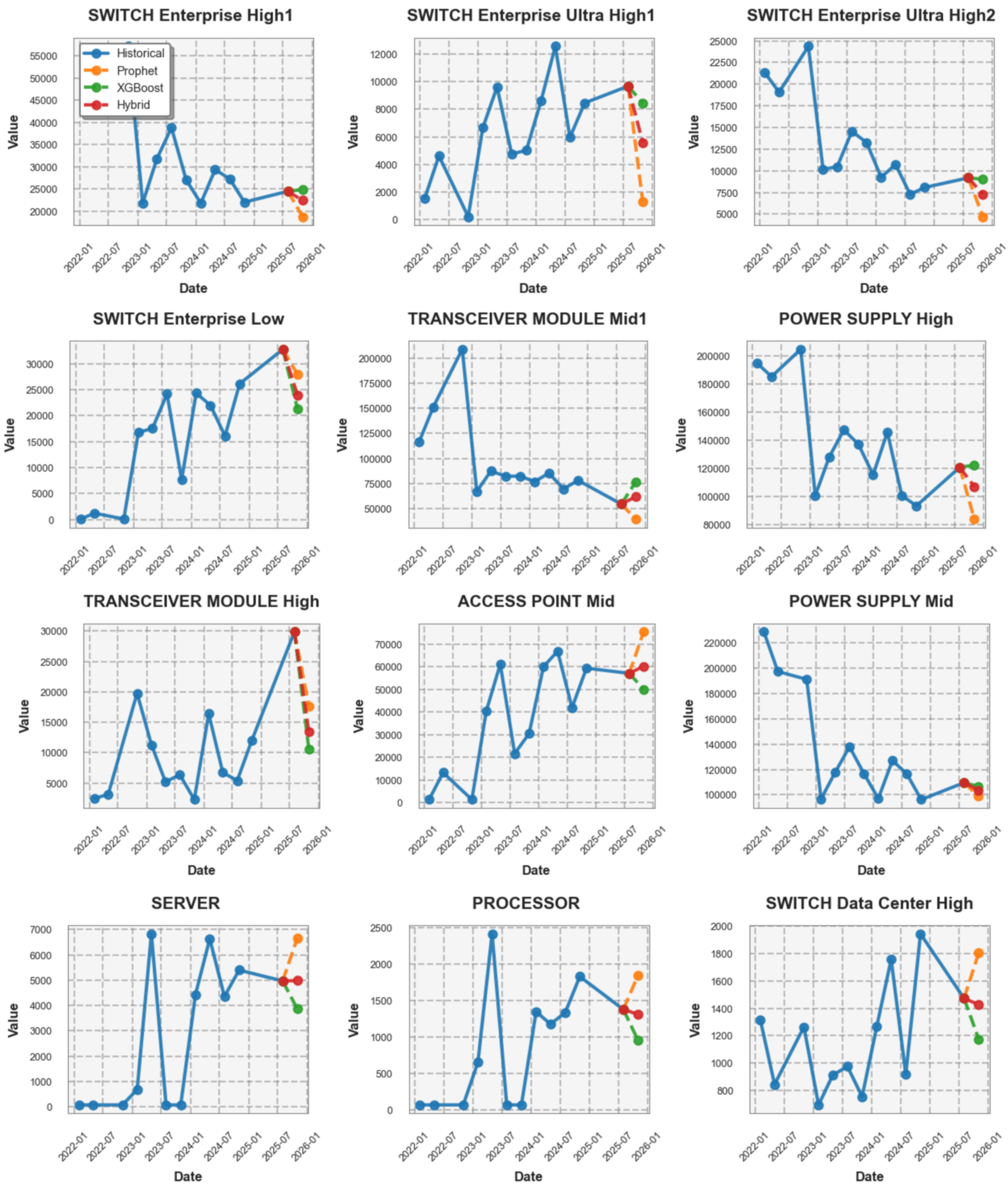
3. Forecasting Trends Visualization

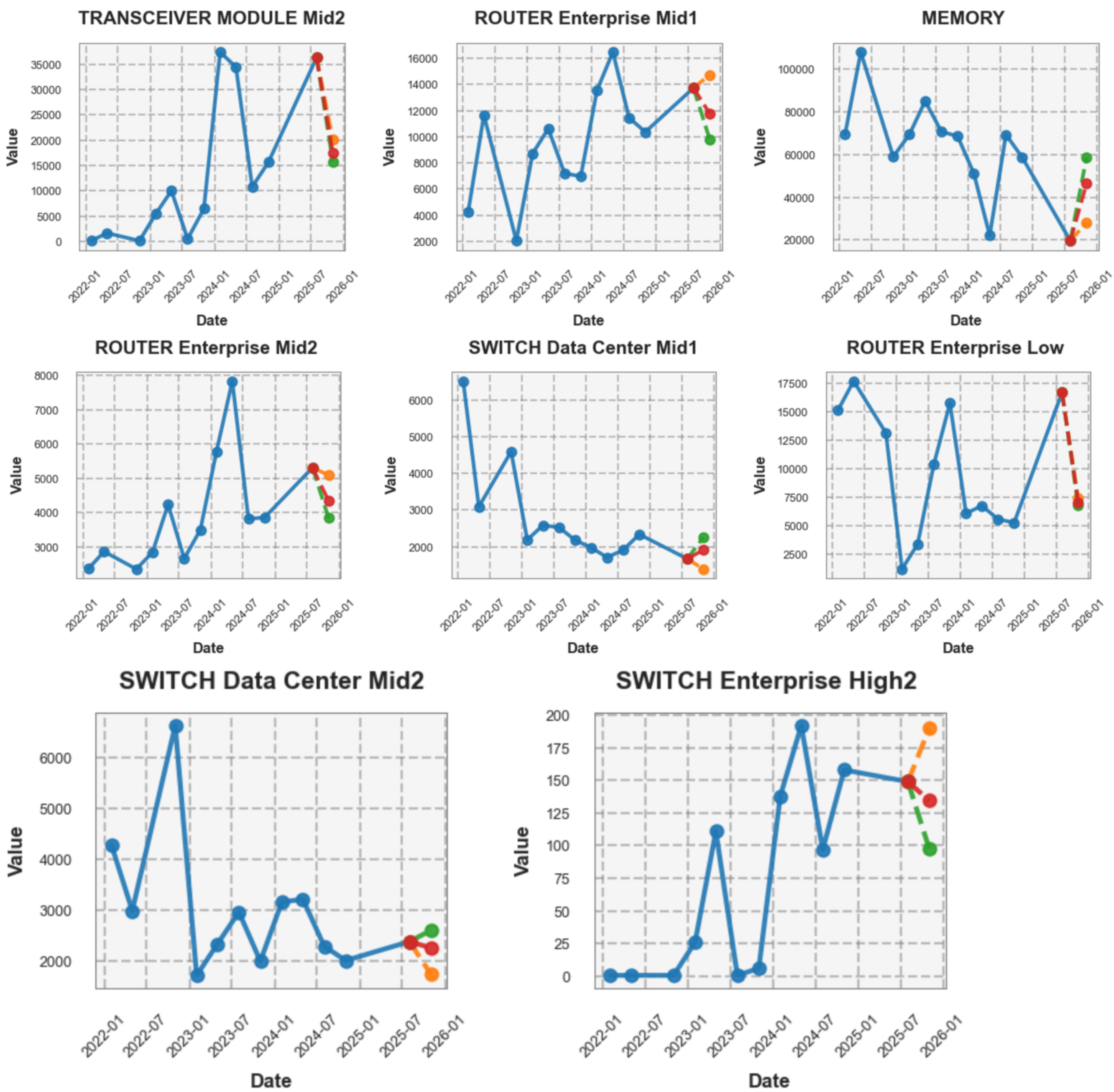
To better understand model performance over time, Figure 1 presents a set of 20 line graphs, each corresponding to a different product. These graphs compare the historical sales data against the predicted values from all three models.

Observations from Forecasting Graphs:

- **XGBoost** performs best for sustaining products, closely following historical trends.
- **Prophet** works well for NPI and declining products but shows inconsistencies elsewhere.
- **Hybrid Model** balances both approaches, reducing errors where individual models struggle.
- **Seasonality Issues**: Prophet assumes seasonal trends, sometimes leading to mispredictions, while XGBoost adapts better to irregular patterns.

- **Sudden Spikes & Declines**: XGBoost captures fluctuations well, Prophet smooths them, and the hybrid model balances both.
- **Confidence & Reliability**: Prophet's forecasts show higher uncertainty, while XGBoost and the hybrid model provide tighter, more reliable predictions.
- **Overall Accuracy**: The hybrid model minimizes errors across diverse product categories, making it the most balanced forecasting approach



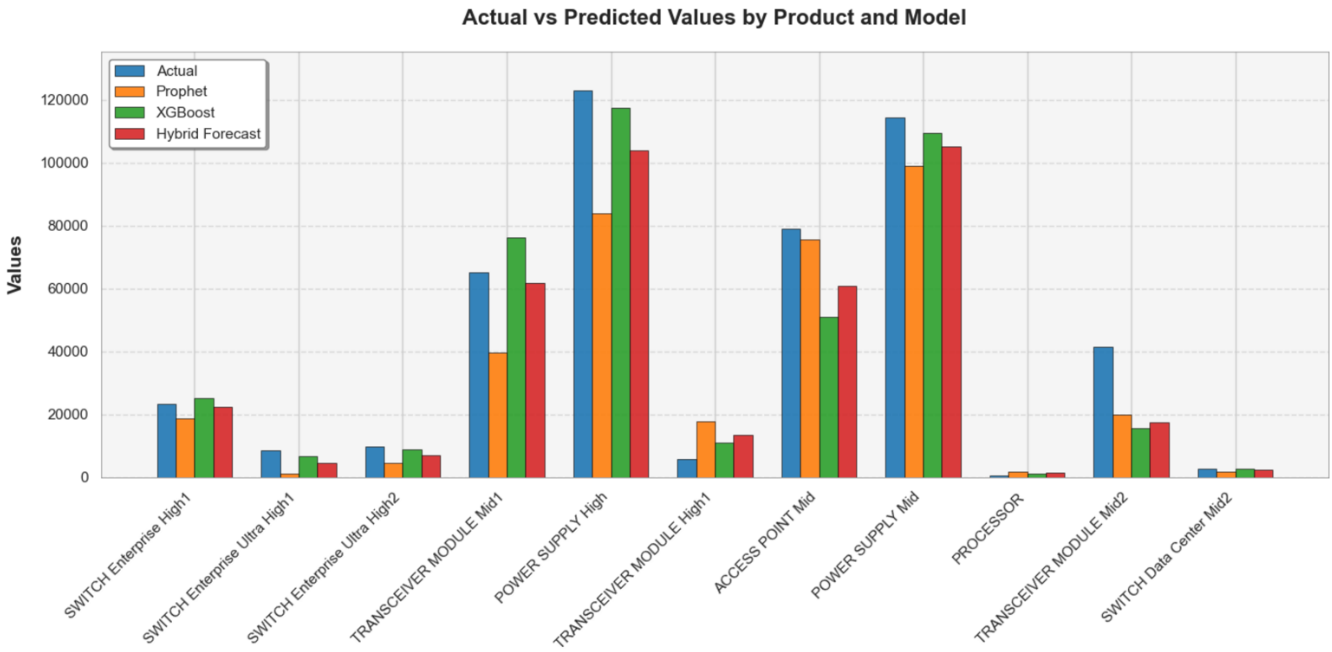


4. Actual vs. Predicted Sales Comparison

A bar graph comparison between actual and predicted sales further validates the models' effectiveness. Figure 2 provides a side-by-side analysis of deviations from actual sales values, helping identify which models consistently provide accurate results.

Insights from Bar Chart Analysis:

- The Hybrid model achieves closer approximations for high-volume products.
- XGBoost tends to slightly underpredict in a few cases but maintains overall accuracy.
- Prophet's overestimation for certain low-volume products is evident.



5. Conclusion

From the above analysis, the Hybrid model emerges as the most robust forecasting approach, balancing the strengths of Prophet and XGBoost. However, in specific stable-demand scenarios, XGBoost proves to be an effective standalone model. The visual comparisons reinforce the necessity of model selection based on product demand patterns.

These insights can aid in optimizing inventory planning and supply chain management by leveraging AI-driven forecasting methodologies tailored to product characteristics. Overall, the combination of multiple forecasting techniques provided a comprehensive view of future demand, allowing for more informed decision-making. The results emphasize the importance of selecting forecasting models based on product characteristics rather than a one-size-fits-all approach. These findings can guide inventory management and resource allocation strategies, leading to improved operational efficiency and reduced forecasting errors. This project is industry ready.