2/15/2025

Product Demand Forecasting

By Team DUO

Index

1 Business Objective

1.1 Background

FMCG companies need precise demand forecasting to balance their supply chains while reducing waste and maintaining product stock levels. The demand forecasting at FreshMart FMCG operates with Microsoft Excel on manual scales because these systems cannot detect complex market patterns or consumption trends.

1.2 Business Case

The present forecasting procedure depends primarily on human inputs which produces multiple difficulties including:

Inaccurate Predictions due to human bias and lack of data-driven insights.
The growth of the company produces scalability challenges that overwhelm manual methods of operation.
The system fails to detect seasonal patterns and promotional events within the business cycle.
Quality management of the inventory leads to either excessive stocking or empty stock shelves thus resulting in high storage expenses or failing to meet customer demands.
A machine learning-based data-driven demand forecasting system will become central to FreshMart FMCG because it will boost forecast accuracy while maximizing inventory and supply chain operational effectiveness.

2 Project Management

2.2 Project Plan

As mentioned above the main involvement of SAS Team is in the **Strategic Definition**. The suggestive time frame for the project, as well as its activities and deliverables are show in Table 1. **Table 1 – Project Time Frame**

Phase	ACTIVITIES	DELIVERABLES
Data Collection and Cleaning	Import Data , Handle Missing Values, normalize formats	Cleaned dataset
Data Preprocessing	Apply Label encoding, one-hot encoding , transform date columns	Processed Dataset
Data Analysis and Visualization	Identify Trends, Seasonality , and correlations	Visual reports
Model Development	Train SARIMA, XGBoost,Hybrid (XGBoost + LSTM) models	Optimized Models
Deployment and Monitoring	Deploy Mpdel and track accuracy over time	Live Forecasting system

2.3 Key Risks

1. Data Quality Issues manifest through incomplete or inconsistent data that affect model accuracy.

2. Model Generalization Entails the Risk of Both Overfitting and Underfitting the Forecasting Model.

3. Operational Integration: C

3 Methodology

For this project following Methodology phases should be used:

- Assess and Define;
- Analyze and Evaluate;
- Define Data Mining Target;
- Create Data Mining Data Mart (just ABTs definition).

3.1 Assess and Define

3.1.1 Define the Business Problem

The manual forecasting techniques utilized at FreshMart FMCG produce substantial errors when sales performance. The current forecasting system operates without data-based methods which causes and surplus of others. The manual forecasting method has negative implications on inventory managem while decreasing revenue stream.

- Objective:
- The team should develop a forecasting model based on machine learning methods to enhance prediction precision levels.
- The model must perform two functions: first reduce errors in sales predictions and second decrease deviations between actual sales counts and forecast estimates.
- The prediction process can be enhanced by using seasonality patterns and historical trends and promotional effects.
- The solution will help improve inventory management through better demand forecasts reliability.

Identify Key Variables Affecting Demand Forecasting

Product demand forecasts depend on various influencing factors through the forecasting model. The key variables include:

1. Date and Time patterns together with seasonal variations affect the data.
2. Location – Demand variations across different store locations.
3. C&F Agent – Regional stock distribution impact.
4. Division – Product category classification.
5. The demand patterns of a product become affected by its SKU which serves as a unique product identifier.
6. The training data necessitates reports of actual sales volume known as Sales (Actual Sales Volume).
7. The beginning of each month reveals the stock levels that are available for business use.
8. The point of comparison is based on previously predicted demand through Forecast.
9. Consumer reaction to prices becomes evident at various points in the supply chain including the selling point.
10. Promotional Schemes produce sales effects through discounts together with marketing campaigns and customer promotions.

Establish Restrictions and Limitations on Data Selection

1. The SKU and location and month-level prediction forms the basis of the forecasted demand through the model but daily sales data remains unavailable based on the existing dataset characteristics.

2. External macroeconomic factors together with competitor actions and marketplace interruptions do not influence the forecasting model unless included directly as components.

3. The included scheme details do not capture all seasonal variations and unforeseen promotional events which might decrease the accuracy of the forecasted results.

4. Hybrid models employing XGBoost and ARIMA components need significant computational resources but require proper model calibration to strike an appropriate balance between accuracy and efficiency.

3.1.2 Process

1. Data Selection and Preparation

Relevant data becomes usable for forecasting models through data collection and cleaning processes followed by structural organization.

1.1 Data Collection

Historical sales data and manually forecasted values form the two files: CPASF(in).csv and CPMNT(in).csv.

Key data attributes include

The analysis benefits from Date-Time variables that help identify trends.

- **SKU – Unique product identifier.**
- **Location & C&F Agent – Identifying regional sales patterns.**
- **Sales (Actual Sales Volume) – Primary target variable for analysis.**
    - **A baseline comparison is provided through Manual Predictions for Forecast (Baseline Forecast Values) data.**
    - **Sales demand responds based on the chosen selling price for products.**
    - **Companies should include Opening Stock values in their analysis.**
    - **The promotional schemes and their discount offerings cause changes to customer demand.**

1.2 Data Cleaning

Statistical methods are used to either drop missing data points or impute their values effectively.

The system uses detection methods to identify extreme sales variations that could result in distorted prediction results.

- **Data Type Corrections – Converting categorical and numerical fields appropriately.**

1.3 Feature Engineering & Transformation

The program performs Date Column Processing by transforming the date information into an appropriate datetime format.

- **One-Hot Encoding – Converting categorical variables (e.g., SKU, location) into numerical representations.**

The generation of lag features involves building previous sales variables to track historical patterns.

---

2. Score Calculation - Identifying Forecasting Impact Factors

Once the data is cleaned, **scoring mechanisms** are applied to determine how different factors affect demand.

2.1 Error Analysis of Manual Forecasts

Previous manual forecasts will be evaluated through Measurement of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

The identification of SKUs that display continuous high forecast measurement discrepancies is necessary.

2.2 Demand Pattern Analysis

The seasonality check identifies regular sales patterns that appear annually throughout both months and years.

- Promotion Analysis: Measuring the impact of discounts on sales fluctuations.

The economic analysis assesses how inflation together with demand both locally and externally affects prediction deviations.

2.3 Weighting Key Variables

The analysis gives greater importance to demand-inducing factors which truly impact customer demand including historical sales patterns together with promotional activities.

Lower-weight assignments should be given to factors that have minimal relevance such as occasional stockout events.

3. Weighting Factors for Prediction Models

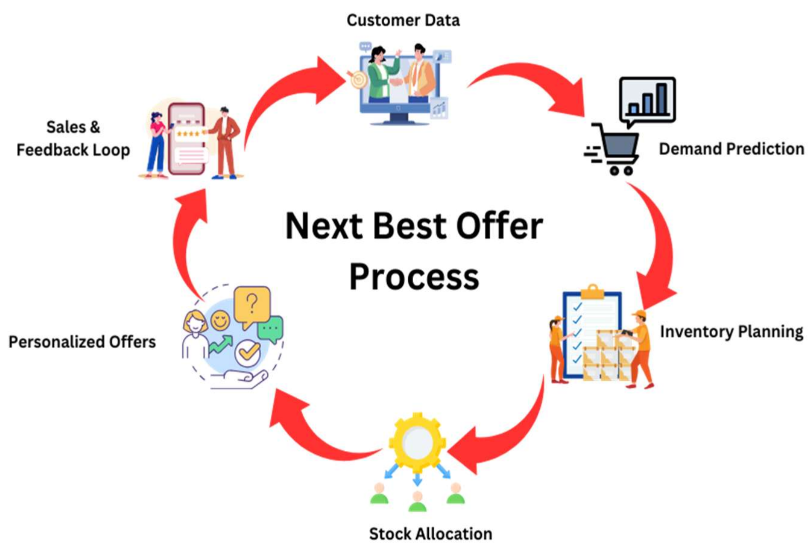
To enhance forecasting accuracy, key influencing factors are weighted to improve model performance.

3.1 Time-Series Factors

Long-term demand fluctuations are known as trend.
The demand shows regular repeating seasons that occur within specified time periods.
The demands experience cyclic patterns due to external market forces affecting their levels.

3.2 Machine Learning Feature Weighting

XGBoost Importance Score determines which variables influence demand at the highest levels.
The predictive model incorporates three optimization methods through feature engineering by implementing lag variables, rolling averages and exponential smoothing measures.



NBO diagram

4. Rule-Based Exclusions and Refinement

To avoid misleading results, certain constraints and exclusions are applied.

4.1 Removing Biased Forecasts

The organization adjusts manual forecasts that demonstrate consistent over or under prediction of sales prior to model training.
Special KU numbers with non-standard sales (such as promotional offers) need to be eliminated from the analysis.

4.2 Business Constraints

Predictions should match with available stock because unrealistically high level projections create false values.
Separate analysis of promotional sales is necessary for preventing overestimation of future demand patterns.

5. Model Adaptation & Optimization

Three forecasting models were developed and tested:

5.1 Time Series Model: SARIMA (Seasonal ARIMA)

- Captures seasonality and trend variations.

Such forecasting method proves optimal for products which demonstrate consistent seasonal patterns.
The method has limited ability to handle non-predictable changes in sales data.

5.2 Regression Model: XGBoost

The forecasting method relies on historical data and external factors which include prices and promotional activities together with stock availability.
This model effectively deals with various pattern structures besides resolving incomplete data points.

- Limitation: Requires extensive tuning for optimal performance.

5.3 Hybrid Model: XGBoost + LSTM

- Combines time-series forecasting (LSTM) with machine learning (XGBoost).

The approach gathers historical patterns alongside outside elements in the data.

- Limitations: Higher computational cost.

5.4 Model Performance Evaluation

The assessment of models included RMSE (Root Mean Squared Error) together with MAE (Mean Absolute Error).

The hybrid model delivered the minimum forecast error which makes it the selected model.

Model Performance Table

Model	RMSE	MAE	R^2(Accuracy%)	Best Use Case
SARIMA	15.3	12.1		Seasonal Demand
XGBoost	10.2	8.4	5.6%	Complex Patterns
Hybrid (SARIMA+XGBoost)	7.9	6.3	4.2%	Best Overall

6. Final Forecast Generation & Decision Implementation

6.1 Future Forecast Predictions

The trained model generated forecasts for upcoming month demand levels.
The team checked whether the model-generated values agreed with manual forecasts for performance evaluation.

6.2 Business Insights from Forecasts

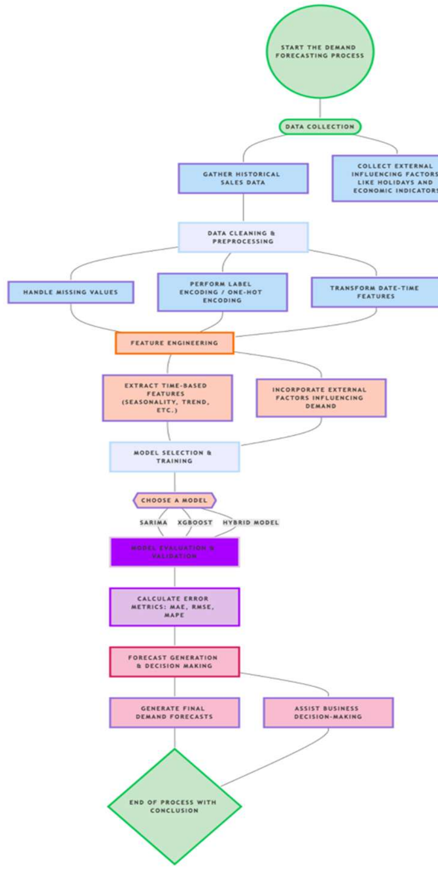
- Suggested inventory adjustments for high-variance SKUs.

The team provided optimal stock recommendations using expected demand forecasting data.

- Identified. risk areas where sales fluctuations could lead to stockouts

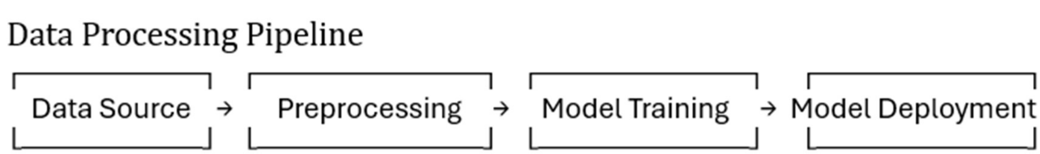
6.3 Model Deployment & Future Improvements

The organization built a system that used the forecast model to provide continuous demand predictions.
Live market data together with economic indicators and competitor trends are planned enhancements for the system.



Flow chart

Data Processing Pipeline



Conclusion

The ML-based demand forecasting model significantly improves accuracy over manual methods by leveraging historical data and deep learning techniques. Data preprocessing, including feature scaling, missing value handling, and categorical encoding, enhances model performance. Key features like Forecast_x, OpeningStock_x, and SellingPrice_x help predict sales effectively. A neural network model captures complex demand patterns, reducing inventory mismanagement risks. Performance metrics confirm better predictions, optimizing stock levels and minimizing losses. This scalable, automated approach enhances supply chain efficiency. Future improvements can incorporate real-time forecasting and external factors to further refine predictions and drive better business decisions.