PROJECT REPORT : AI-Powered Demand Forecasting for Products

Submitted by : Adityabaan Tripathy
ID.No.: IITRPRAI_24091312

Problem Statement:

Stockpiling appropriate quantities is extremely important to the smooth running of a business. Demand forecasting was generally considered an art form, and any imprecise estimation could result in stockouts, loss of business opportunity, and diminished prestige in the eyes of its consumers. Unlike a single factor demand forecast that typically considers past sales figures, multi-factor demand meetings will be held to address issues such as marketing campaigns, seasonal trends, pricing, actions taken by competitors, and stock levels. This project will develop an AI-driven demand forecasting system that incorporates these factors through machine learning models to provide demand predictions for retail products that are both accurate and actionable.

Methodology:

The dataset used consists of historical sales records, including product IDs, base sales, marketing campaign types, seasonal trends, prices, discounts, competitor prices, stock availability, public holidays, and actual demand values, covering multiple years on a weekly basis.

Data Preprocessing:

- Date conversion and extraction of year and week features for temporal structuring.
- Aggregation of demand by year, week, product, and base sales groups.
- Encoding of categorical variables for Product_ID and Base_Sales into numerical codes for modeling.
- Creation of lag features (demand in previous weeks) to capture temporal dependencies.
- Addition of a quarter feature to model seasonality.

Model Development:

Three machine learning models were developed and compared: Random Forest, Gradient Boosting, and XGBoost regressors.

- Dataset was split into training (80%) and test (20%) sets.
- Models were trained on features—year, week, product code, store code, lag features, and quarter.
- Performance metrics Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were calculated on test predictions.

Tools & Technologies:

- Programming Language: Python 3.
- Development Environment: Google Colab for cloud-based execution and easy sharing
- Libraries: pandas and numpy for data manipulation, matplotlib and seaborn for visualization, scikit-learn for machine learning models, xgboost for advanced boosting, and shap for model explainability.
- Dataset: Publicly provided sales data with multiple influencing features across time.

Results:

Model	RMSE	MAE
Random Forest	44,290.27	32,506.80
Gradient Boosting	43,185.97	31,442.08
XGBoost	43,570.65	31,748.87

- Gradient Boosting was the best performer and showed a lower value of RMSE and MAE, indicating a more accurate forecast. Comparing actual and forecasted sales of top products shows a close match, thus validating the model.
- Feature importance of the Random Forest model showed lag features and quarter to be significant, emphasizing the temporal and seasonal effects on demand. SHAP explainability plots gave a deeper insight into feature contributions in individual predictions.
- The notion of seasonality went further with average sales reaching peaks at certain quarters, justifying the need to create season-aware models.

Conclusion:

This project established a successful AI-driven demand forecasting framework where various influencing factors are integrated into the preparation of accurate product demand forecasts. After comparing various models, it was concluded that ensemble-based methods are viable and win better at capturing complex demand behavior. The feature importance and explainability provide actionable business insights relating to inventory planning and marketing strategies. Limitations pertain to historical data quality, and the use of no real-time external data streams such as social media sentiment. The future could extend the model toward real-time data, deeper time-specific series models such as LSTM, and implement it as an interactive dashboard to assist with dynamic business decision-making.

Dataset : Product Demand Forecasting Dataset (Kaggle)

URL : https://www.kaggle.com/datasets/chavindudulaj/product-demand-forecasting-dataset?resource=download