**Dominos - Predictive Purchase Order System**

**Domain Introduction:**  
The Food Service Industry is characterized by high variability in demand and a complex supply chain. Accurate sales forecasting is crucial for inventory management and minimizing waste. Predictive modeling can streamline operations and improve efficiency.

**Problem Statement:**  
Optimize the process of ordering ingredients by predicting future sales and creating a purchase order.

**Data Cleaning and Preprocessing:**  
Techniques used include null value imputation with mean/mode, handling outliers using IQR method, and data normalization to ensure consistency and accuracy.

**EDA:**

* **Univariate Analysis:** Identified sales trends and seasonality.
* **Bivariate Analysis:** Examined relationships between sales and factors like day of the week and promotional periods.
* **Multivariate Analysis:** Combined insights to identify key sales drivers.

**Feature Engineering:**  
Generated features such as day of the week, month, and holiday periods to enhance model accuracy.

**Statistical Significance:**  
Used the Augmented Dickey-Fuller test to check for stationarity in time series data, ensuring the validity of our forecasting models.

**Class Imbalance Technique:**  
Applied SMOTE to handle any imbalances in categorical features related to promotional periods or holidays.

**Model Building:**

* **Base Model:** Started with ARIMA due to its effectiveness in time series forecasting.
* **Additional Models:** Explored SARIMA, Prophet, and regression model XGBoosting to capture different aspects of seasonality and trend. Hyperparameter tuning was performed using grid search.

**Model Evaluation Metric:**  
Mean Absolute Percentage Error (MAPE) was chosen for its interpretability and relevance to business impact.

**Final Model:**  
Selected **Prophet** as the final model due to its superior performance in capturing seasonality and trend patterns.

**Conclusion:**  
Feature importance analysis highlighted the significance of promotional periods and holidays in driving sales.

**Business Suggestion/Solution:**  
Implement the predictive purchase order system to reduce waste, optimize inventory levels, and align supply chain operations with forecasted demand, ultimately enhancing operational efficiency and customer satisfaction.

**Detailed Explanation of ARIMA, SARIMA, Prophet, and XGBoost Models**

**ARIMA (AutoRegressive Integrated Moving Average)**

**Introduction:** ARIMA is a widely used time series forecasting method that combines three components: autoregression (AR), differencing (I), and moving average (MA). It is suitable for univariate time series data.

**Components:**

* **AutoRegressive (AR):** Refers to the model's dependency on its own previous values. It uses a specified number of lagged observations to model the relationship.
* **Integrated (I):** Represents the differencing of raw observations to make the time series stationary. Differencing helps to remove trends and seasonality.
* **Moving Average (MA):** Uses past forecast errors in a regression-like model.

**Mathematical Representation:** Yt=c+∑i=1pϕiYt−i+∑j=1qθjϵt−j+ϵtY\_t = c + \sum\_{i=1}^{p} \phi\_i Y\_{t-i} + \sum\_{j=1}^{q} \theta\_j \epsilon\_{t-j} + \epsilon\_tYt​=c+∑i=1p​ϕi​Yt−i​+∑j=1q​θj​ϵt−j​+ϵt​ Where YtY\_tYt​ is the forecasted value, ccc is a constant, ϕi\phi\_iϕi​ are the autoregressive coefficients, θj\theta\_jθj​ are the moving average coefficients, and ϵt\epsilon\_tϵt​ is the error term.

**Strengths:**

* Good for short-term forecasting.
* Captures linear relationships effectively.

**Limitations:**

* Not suitable for data with complex seasonal patterns.
* Requires the time series to be stationary.

**SARIMA (Seasonal ARIMA)**

**Introduction:** SARIMA extends ARIMA to support seasonal data. It includes additional seasonal components to account for patterns repeating at regular intervals.

**Components:**

* **Seasonal AutoRegressive (SAR):** Similar to AR but for seasonal lags.
* **Seasonal Integrated (SI):** Similar to I but for seasonal differencing.
* **Seasonal Moving Average (SMA):** Similar to MA but for seasonal lags.

**Mathematical Representation:** Yt=c+∑i=1pϕiYt−i+∑j=1qθjϵt−j+∑k=1PΦkYt−km+∑l=1QΘlϵt−lm+ϵtY\_t = c + \sum\_{i=1}^{p} \phi\_i Y\_{t-i} + \sum\_{j=1}^{q} \theta\_j \epsilon\_{t-j} + \sum\_{k=1}^{P} \Phi\_k Y\_{t-km} + \sum\_{l=1}^{Q} \Theta\_l \epsilon\_{t-lm} + \epsilon\_tYt​=c+∑i=1p​ϕi​Yt−i​+∑j=1q​θj​ϵt−j​+∑k=1P​Φk​Yt−km​+∑l=1Q​Θl​ϵt−lm​+ϵt​ Where Φk\Phi\_kΦk​ and Θl\Theta\_lΘl​ are the seasonal autoregressive and moving average coefficients, respectively, and mmm is the seasonal period.

**Strengths:**

* Handles seasonality effectively.
* Suitable for both short and long-term forecasting.

**Limitations:**

* More complex and computationally intensive than ARIMA.
* Requires more parameters to be tuned.

**Prophet**

**Introduction:** Prophet, developed by Facebook, is a robust and flexible forecasting tool designed for business time series data. It handles missing data and seasonal trends efficiently.

**Components:**

* **Trend:** Piecewise linear or logistic growth curves.
* **Seasonality:** Yearly, weekly, and daily seasonality modeled using Fourier series.
* **Holidays:** Special handling for user-specified holiday effects.

**Mathematical Representation:** Y(t)=g(t)+s(t)+h(t)+ϵtY(t) = g(t) + s(t) + h(t) + \epsilon\_tY(t)=g(t)+s(t)+h(t)+ϵt​ Where g(t)g(t)g(t) is the trend function, s(t)s(t)s(t) is the seasonality function, h(t)h(t)h(t) represents holiday effects, and ϵt\epsilon\_tϵt​ is the error term.

**Strengths:**

* Handles outliers and missing data well.
* Flexible in incorporating seasonal and holiday effects.
* User-friendly with intuitive parameters.

**Limitations:**

* May require domain knowledge to set appropriate seasonalities and holidays.
* Performance may vary with different types of time series data.

**XGBoost (Extreme Gradient Boosting)**

**Introduction:** XGBoost is a powerful and efficient implementation of gradient boosting for supervised learning tasks, including regression and classification. It is known for its high performance and accuracy.

**Components:**

* **Gradient Boosting:** An ensemble technique that builds models sequentially, each new model correcting errors made by the previous ones.
* **Decision Trees:** Used as the base learners, capturing non-linear relationships.

**Mathematical Representation:** yi=∑k=1Kfk(xi)y\_i = \sum\_{k=1}^{K} f\_k(x\_i)yi​=∑k=1K​fk​(xi​) Where yiy\_iyi​ is the predicted value, fkf\_kfk​ represents the individual trees, and KKK is the number of trees.

**Strengths:**

* High predictive power.
* Handles missing data and outliers effectively.
* Feature importance ranking.
* Supports regularization to prevent overfitting.

**Limitations:**

* Computationally intensive.
* Requires careful parameter tuning.
* May be prone to overfitting if not properly regularized.

**Choosing the Models**

**ARIMA:** Chosen for its simplicity and effectiveness in handling short-term linear relationships in time series data.

**SARIMA:** Extended from ARIMA to handle seasonality, making it suitable for data with regular seasonal patterns.

**Prophet:** Selected for its flexibility in dealing with missing data, outliers, and complex seasonal effects. Its user-friendly interface allows easy customization for holidays and special events.

**XGBoost:** Used for its high performance in predictive modeling tasks. Although primarily a machine learning model for tabular data, it can be adapted for time series forecasting by creating lag features and capturing complex patterns in the data.

**Conclusion**

Each model brings unique strengths to time series forecasting and predictive modeling. ARIMA and SARIMA are preferred for their simplicity and ability to handle linear relationships and seasonality. Prophet is highly flexible and user-friendly, suitable for complex seasonal patterns and missing data. XGBoost offers high predictive power and robustness, making it ideal for capturing non-linear relationships in the data. Combining these models allows leveraging their strengths to achieve accurate and reliable forecasts, ultimately aiding in better inventory management and business decision-making.