R&D Document for Time Series Forecasting

1 Introduction

In this task, I aim to build a time series forecasting model that predicts the number of units sold for each item ID based on sales data from a well-known brand on Amazon. The primary evaluation metric is the Mean Squared Error (MSE).

2 Data Understanding

The dataset consists of the following columns:

- **date**: The date of the sales data (YYYY-MM-DD).
- **Item Id**: A unique identifier for each item.
- **Item Name**: The name of the item.
- **anarix_id**: An internal identifier for tracking items.
- **ad_spend**: The advertising spend on the given date.
- **units**: The number of units sold (target variable).
- **orderedrevenueamount**: The total revenue generated from sales.
- **unit_price**: The price per unit.

3 Exploratory Data Analysis (EDA)

EDA helps in understanding the data distribution and identifying patterns. Key steps include:

- Checking for missing values.
- Plotting sales trends over time.
- Visualizing the relationship between ad spend and sales.

4 Feature Engineering

Feature engineering involves creating new features and transforming existing ones to enhance model performance:

- Convert **date** to datetime format.
- Extract **year**, **month**, **day**, and **day of the week** from the date.
- Handle missing values (e.g., filling missing **ad_spend** with 0).

5 Model Selection

I chose the Facebook Prophet model for time series forecasting due to its advantages:

- Handles missing data and outliers.
- Supports daily seasonality and holiday effects.
- Provides interpretable model components (trend, seasonality, holidays).

6 Hyperparameter Tuning

Hyperparameter tuning involves adjusting model parameters to improve performance. For Prophet, we tuned parameters like **changepoint_prior_scale** and **seasonality_prior_scale**.

7 Model Evaluation

Model evaluation is performed using the Mean Squared Error (MSE), which measures the average squared difference between predicted and actual values. Lower MSE values indicate better model performance.

8 Conclusion

Based on my analysis and modeling, I forecast the number of units sold for each item ID. The detailed steps and code implementation are provided in the attached notebook.