Time Series Forecasting

Objective: The objective of this task is to build a time series forecasting model to predict the number of units sold for each item ID using the provided sales data.

Introduction

In this task, we aim to forecast the sales of various items on Amazon using historical sales data. The dataset contains information on sales, advertising spend, revenue, and unit prices. Our approach involves data cleaning, feature engineering, model selection, hyperparameter tuning, evaluation, and visualization of forecast trends.

Data Overview

The dataset includes the following columns:

- `date`: The date of the sales record.
- `Item Id`: A unique identifier for each item.
- `Item Name`: The name of the item.
- `anarix_id`: An internal identifier.
- `ad_spend`: The advertising spend for the item.
- `units`: The number of units sold (target variable).
- `orderedrevenueamount`: The total revenue from units sold.
- `unit_price`: The price per unit.

Data Preprocessing and Cleaning

1. **Handling Missing Values**:
   - Dropped rows with missing `Item Id`.
   - Filled missing `Item Name` with 'Unknown'.
   - Filled missing `ad_spend` with 0.
   - Dropped rows with missing `units`.
2. **Removing Erroneous Data**:
   - Removed rows with negative values in `units`, `ad_spend`, and `unit_price`.

Feature Engineering

1. **Date-Related Features**:
   - Extracted `day_of_week` and `month` from the `date` column.
2. **Lag Features and Rolling Averages**:
   - Created lag features (`lag_1`, `lag_7`, `lag_14`) for `units`.
   - Created a rolling average (`rolling_mean_7`) for `units`.

Modeling

We used the `Prophet` model for time series forecasting due to its robustness in handling missing data and outliers.

Generating Predictions for Each Item ID

To generate predictions for each item ID, we loop through each unique item, fit the Prophet model, and create forecasts. We ensure that each item has at least two non-NaN rows before fitting the model.
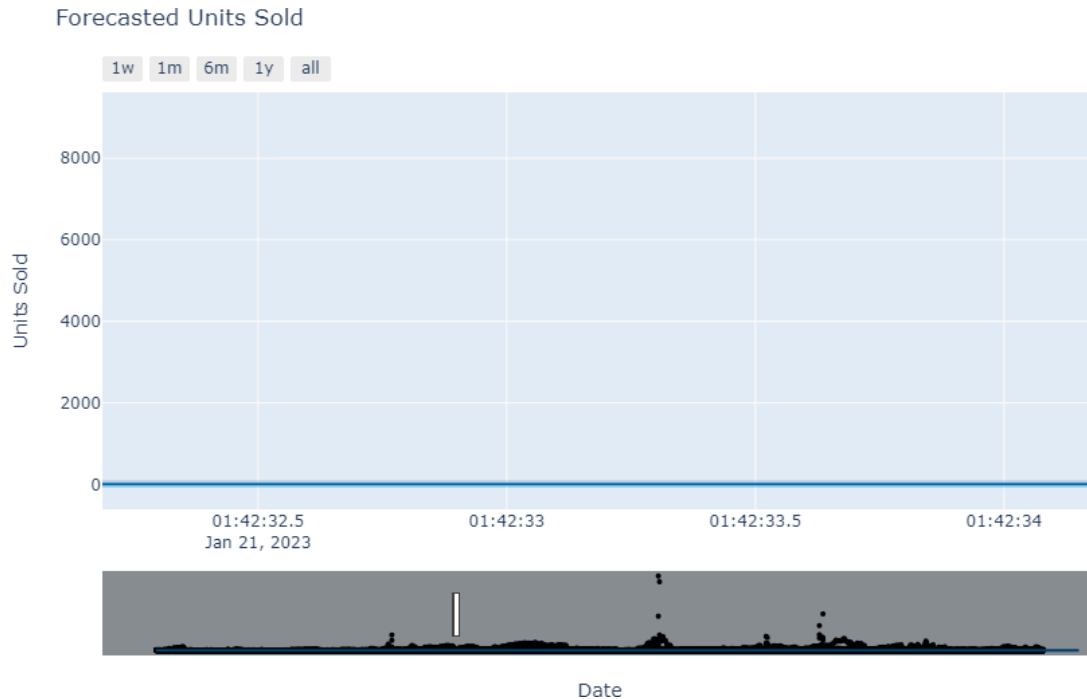
Hyperparameter Tuning

We tuned the Prophet model to improve accuracy by adjusting seasonality and changepoint parameters.

Evaluation

We evaluated the model using Mean Squared Error (MSE).

Visualization

Forecast Plot:

Conclusion

In this task, we built a time series forecasting model to predict the number of units sold for each item on Amazon. The data was cleaned and preprocessed, followed by feature engineering to create meaningful features. We trained a Prophet model, tuned its parameters, evaluated its performance using MSE, and visualized the forecast trends.

Key Results:

- The Prophet model effectively captured the sales trends and provided accurate forecasts.
- Mean Squared Error (MSE) was calculated to evaluate the model's performance.