Time Series Forecasting Project

Name - Manoj Kumar Sen
Reg.no. - 20MIP10022
Contact – 9926200426
College – VIT Bhopal

Objective

The objective of this project is to build a time series forecasting model to predict the number of units sold for each item ID using dummy sales data from a well-known brand on Amazon. The primary evaluation metric for the model's performance is the Mean Squared Error (MSE).

1. Loading the Data

We began the project by loading the dataset into our working environment. The dataset contained columns such as:

date: The date of the sales transaction.
Item Id: A unique identifier for each item.
units: The number of units sold on that date for that item.
anarix_id: An internal identifier that might be used for tracking or categorizing items within the brand's system.
ad_spend: The amount of money spent on advertising for the item on the given date. This column can be useful in understanding the relationship between advertising spend and sales performance.

2. Data Preprocessing

To ensure the data was ready for analysis, we performed several preprocessing steps:

Conversion of Date Column: We converted the date column to a datetime format to facilitate time series analysis.
Filtering Data: We filtered the dataset to focus on specific item IDs to perform detailed analysis and forecasting.
Setting Index: The date column was set as the index of the dataframe to align with time series analysis requirements.

3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to understand the underlying patterns and characteristics of the data. Key steps included:

Visualization of Sales Data: We plotted the sales data to visualize overall trends and seasonal patterns. This helped in identifying any obvious outliers or irregularities.
Innovative Visualizations: We created various plots such as line charts and subplots to highlight daily, weekly, and monthly seasonality. These visualizations provided deeper insights into sales patterns over different time periods.
Rolling Mean Calculations: We calculated rolling means (e.g., 7-day and 30-day) to smooth out short-term fluctuations and highlight longer-term trends. This helped in better understanding the data’s structure.

4. Feature Engineering
Feature engineering was essential to enhance the predictive power of our model. Steps included:

Extracting Day of the Week: We extracted the day of the week from the date column to analyze daily seasonality patterns.
Calculating Rolling Means: We calculated 7-day and 30-day rolling means of the units sold to capture short-term and long-term trends. These features were used to provide additional context to the model.

5. Model Selection
For model selection, we opted for Facebook's Prophet model due to its effectiveness in time series forecasting and its ability to handle seasonality and trend analysis. Prophet is particularly suitable for business time series data, providing robust and reliable forecasts.

6. Model Training
The dataset was split into training and test sets to evaluate the model's performance:

Training Set: The majority of the data was used to train the model.
Test Set: The last 30 days of data were reserved as the test set to evaluate the model's forecasting accuracy.
We trained the Prophet model on the training data, allowing it to capture underlying patterns, seasonality, and trends in the sales data.

7. Model Prediction
Using the trained Prophet model, we made predictions for the test set period:

Future Dataframe Creation: We created a future dataframe for the next 30 days.
Forecasting: The model generated predictions for the future dataframe, providing the expected number of units sold for each day in the test period.

8. Model Evaluation
To evaluate the model's performance, we calculated the Mean Squared Error (MSE):

Comparison of Actual and Predicted Values: We compared the actual sales data with the predicted values from the model.
Calculation of MSE: The Mean Squared Error was calculated to measure the average magnitude of the errors between the predicted and actual values. A lower MSE value indicated better model performance.

9. Visualization of Results
We visualized the actual versus predicted sales values using line charts:

Plotting: We plotted the actual sales data against the model's predictions to provide a clear visual comparison.
Analysis of Trends and Seasonality: The plot helped in understanding how well the model captured the trends and seasonality in the sales data, allowing us to visually inspect the accuracy of the forecasts.

10. Predicted Result

Linear Regression
Mean Squared Error: 1.759550849674442

ARIMA model
Mean Squared Error: 3.0997320449515633

Prophet
Mean Squared Error: 2.0197320449515633