Parcel Volume Forecasting

1. Introduction

In this report, we analyze the parcel volume data to forecast future trends using multiple machine learning techniques including Linear Regression, ARIMA, Random Forest Regressor, and Prophet. The data used spans several years, and the goal is to provide a reliable forecast based on the model performances.

2. Data Overview

The dataset contains parcel volume data over time. The dataset was preprocessed and time-indexed to make it suitable for time-series forecasting. Date values were converted to ordinal integers for models requiring numerical input.

3. Modeling Techniques

3.1 Linear Regression

Linear Regression is a basic and interpretable model that attempts to capture the linear relationship between the time index and the parcel volume. This model is straightforward, but may not capture complex patterns in time-series data.

3.2 ARIMA

ARIMA is a popular statistical method for time-series forecasting that captures both trend and seasonality. In this case, we used an ARIMA(2,0,2) model to capture the relationships in the parcel volume data.

3.3 Random Forest Regressor

The Random Forest Regressor is an ensemble learning method that can capture complex non-linear relationships. By using multiple decision trees, it can provide robust predictions, especially in diverse datasets.

3.4 Prophet

Prophet is a forecasting tool designed for business time-series data. It is robust to missing data and shifts in the trend, and it automatically detects seasonal patterns.

4. Model Evaluation

The performance of the models was evaluated using Mean Squared Error (MSE) and Mean Absolute Error (MAE). The results are as follows:

Linear Regression

MSE: 52731.59, MAE: 185.50

ARIMA

MSE: 38408.63, MAE: 172.25

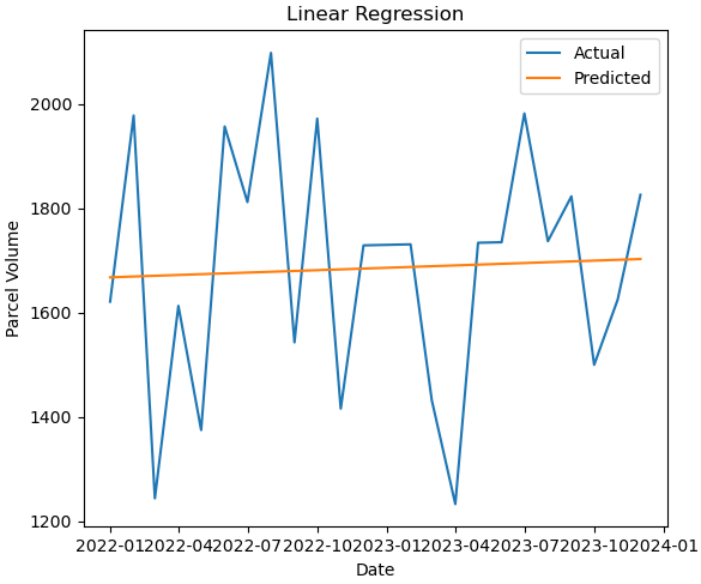
Random Forest Regressor

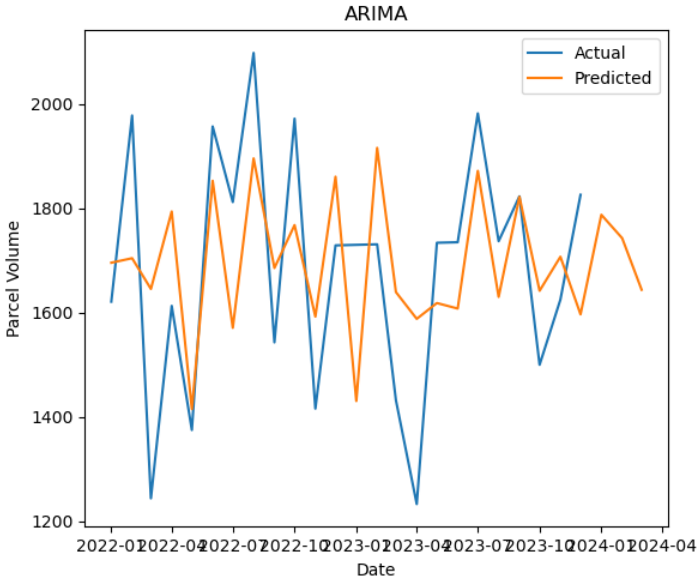
MSE: 13728.82, MAE: 100.29

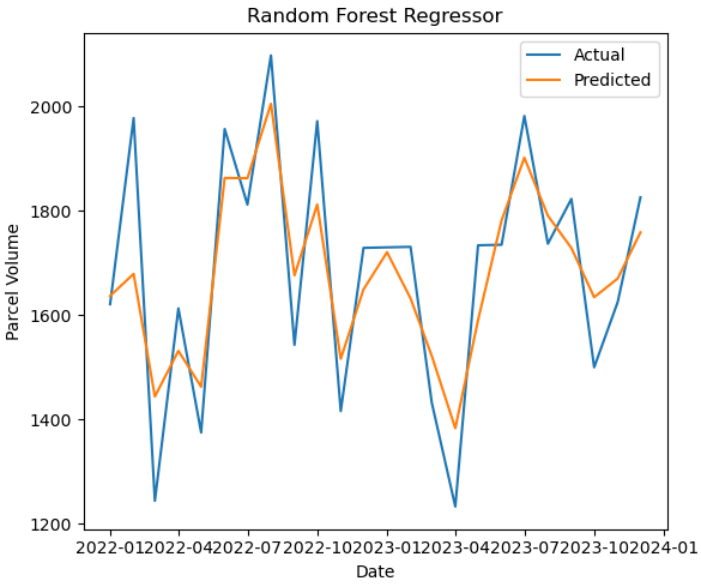
Prophet

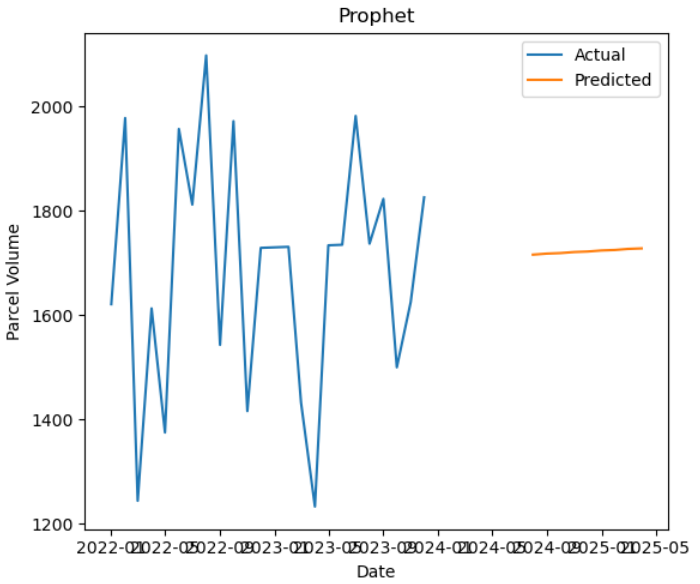
MSE: 42509.00, MAE: 146.11

Forecasting Python









5. Conclusion

Based on the evaluation metrics, we observe that each model has its strengths and weaknesses. The Random Forest Regressor shows promise in capturing non-linear relationships, while ARIMA provides a good fit for time series with trend and seasonality. The choice of model may depend on the specific requirements of accuracy, interpretability, and computational resources.