TSA Forecast Assessment Report
Name: Brighton Mao
Project Manager: Eubene In
Date: February 11, 2026

Methodology:

I created a class called GradientBoostingRegressor which implements a gradient boosting algorithm where the base learner in the DecisionTreeRegressor from the sklearn.tree package. Using the 'Date' column from tsa_train.csv, I added the following numeric columns which became features in my model:

- year
- month
- day
- day_of_week
- is_weekend
- day_of_year
- quarter
- is_holiday

I used a learning rate of 0.01 and 2000 estimators.

Data Analysis:

Since the dataset only contains 'Date' and 'Volume', I knew I had to add more columns/features for my model to learn well. As I began adding more features, my test score kept increasing, so I knew the features were helpful. However, I'd like to know exactly which features are most helpful in this list. If I had more time, I would've used multiple regression and ANOVA to pinpoint the most impactful variables.

Learnings:

This was my first time creating a tree-based model. The most challenging aspect of building this model was understanding the overall structure of the model and how each line of code contributed to its 'learning' . As I look over this model more and create new ones in the future, I hope to better understand tree-based models and learn how to make them more efficient and accurate.