Model Selection and Evaluation Report

Objective

This report evaluates the performance of different machine learning models using two different feature sets:

1. ['Global_intensity', 'Sub_metering_3']
2. ['Global_intensity']

The goal is to determine which model performs best based on accuracy metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-Squared (R^2).

Feature Set 1: ['Global_intensity', 'Sub_metering_3']

1. Linear Regression
   - RMSE: 0.0467
   - MAE: 0.0313
   - R^2: 0.9980

2. Random Forest
   - RMSE: 0.0610
   - MAE: 0.0407
   - R^2: 0.9966

3. Gradient Boosting
   - RMSE: 0.0436
   - MAE: 0.0279
   - R^2: 0.9983

4. Neural Network
   - RMSE: 0.0461
   - MAE: 0.0285
   - R^2: 0.9981

Feature Set 2: ['Global_intensity']

1. Linear Regression

- RMSE: 0.0499
- MAE: 0.0341
- R^2: 0.9978

2. Random Forest
- RMSE: 0.0605
- MAE: 0.0403
- R^2: 0.9967

3. Gradient Boosting
- RMSE: 0.0457
- MAE: 0.0291
- R^2: 0.9981

4. Neural Network
- RMSE: 0.0457
- MAE: 0.0298
- R^2: 0.9981

Analysis and Conclusion

1. Best Performing Model:
- Gradient Boosting achieved the lowest RMSE (0.0436) and MAE (0.0279) with Feature Set 1, making it the most accurate model.
- Neural Network also performed well, but slightly lagged behind Gradient Boosting.

2. Effect of Feature Selection:
- Including 'Sub_metering_3' improved the overall model performance, as Feature Set 1 resulted in lower RMSE and MAE values compared to Feature Set 2.

3. Recommendation:
- Gradient Boosting with Feature Set 1 ('Global_intensity', 'Sub_metering_3') is the best choice for this problem as it yields the highest accuracy and lowest error rates.