Model Evaluation Report: Regression Analysis

1. Introduction

This report presents a comprehensive evaluation of multiple regression models applied to a dataset. The goal is to identify the best-performing model based on various regression metrics. The models evaluated include:

- Linear Regression (3 variations)
- Random Forest Regression (3 variations)
- Gradient Boosting Regression (3 variations)
- Neural Network Regression (3 variations)

2. Evaluation Metrics

Each model is assessed using the following metrics:

- **MAE (Mean Absolute Error):** Average of absolute differences between predicted and actual values.
- **MSE (Mean Squared Error):** Average of squared differences between predicted and actual values.
- **RMSE (Root Mean Squared Error):** Square root of MSE, representing error magnitude.
- **R^2 Score (Coefficient of Determination):** Indicates how well the model explains the variance in the target variable (1.0 is perfect).

3. Performance Summary Table

Model	MAE	MSE	RMSE	R^2 Score
Linear Regression MODEL 1	0.0274	0.00179	0.0423	0.99840
Linear Regression MODEL 2	0.1078	0.02875	0.1695	0.99855
Linear Regression MODEL 3	0.1122	0.03055	0.1748	0.99846
Random Forest MODEL 1	**0.0156**	**0.00096**	**0.0309**	**0.99915**
Random Forest MODEL 2	0.0537	0.01557	0.1248	0.99921
Random Forest MODEL 3	0.0508	0.01450	0.1204	**0.99927**
Gradient Boosting MODEL 1	0.0209	0.00119	0.0345	0.99894
Gradient Boosting MODEL 2	0.0875	0.02060	0.1435	0.99896
Gradient Boosting MODEL 3	0.0874	0.02042	0.1429	0.99897
Neural Network MODEL 1	0.0602	0.00508	0.0713	0.99547
Neural Network MODEL 2	0.1514	0.03493	0.1869	0.99824
Neural Network MODEL 3	0.0860	0.02033	0.1426	0.99897

Model 1:
X = df.drop(columns['Global_active_power','Sub_metering_3'])
y = df['Global_active_power']

Model 2:
X=df.drop(columns['Global_intensity','Sub_metering_3'])
y= df['Global_intensity']

Model 3:
X= df.drop(columns=['Global_intensity'])
y= df['Global_intensity']

4. Analysis

Linear Regression Models

- **MODEL 1** performs best among linear models with low MAE and RMSE.
- However, it is outperformed by tree-based models in all metrics.

Random Forest Models

- **MODEL 1** achieves the **lowest MAE (0.0156)** and **lowest RMSE (0.0309)**.
- It also has a **very high R^2 score (0.99915)**, making it the most reliable overall.
- **MODEL 3** has the highest R^2 (0.99927), but slightly higher error rates than MODEL 1.

Gradient Boosting Models

- These models also perform well, with MODEL 1 achieving competitive results (MAE: 0.0209).
- However, they are slightly less accurate than Random Forests.

Neural Network Models

- These show more variability.
- MODEL 3 performs the best among them (MAE: 0.0860), but it does not outperform the best tree-based models.

5. Final Thoughts

Tree-based models like **Random Forests** and **Gradient Boosting** consistently outperform linear and neural network models in this analysis. Particularly, **Random Forest MODEL 1** stands out due to its superior accuracy and generalization.

Conclusion

After evaluating multiple regression models using standard performance metrics (MAE, MSE, RMSE, and R^2 Score), it is evident that **Random Forest Regression MODEL 1** delivers the most accurate and reliable results. It achieves the **lowest error rates** and a **high R^2 Score of 0.99915**, indicating it explains nearly all the variance in the target variable. While other models such as Gradient Boosting and Neural Networks also performed well, they were slightly less consistent in terms of error minimization.

Therefore, **Random Forest Regression MODEL 1 is recommended as the optimal model** for this task due to its exceptional balance between prediction accuracy and model robustness. It is well-suited for real-world deployment where both precision and reliability are critical.