Shivansh Yadav 102203791 3CO18

Model Performance Report

Dataset Summary

The dataset contains **1,000 samples** and **15 features**, with a target variable named **"Target2"**. The data was split into:

- **Training Set**: 70% (700 samples)
- **Testing Set**: 30% (300 samples)

Models Evaluated

The following regression models were trained and evaluated:

1. Linear Regression
2. Random Forest Regressor
3. Gradient Boosting Regressor
4. Decision Tree Regressor
5. K-Nearest Neighbors Regressor
6. Support Vector Regressor

Model Performance Metrics:

Linear Regression:

Mean Squared Error: 0.1062

Mean Absolute Error: 0.2683

R^2 Score: 0.6356

Random Forest Regressor:

Mean Squared Error: 0.1116

Mean Absolute Error: 0.2746

R^2 Score: 0.6169

Gradient Boosting Regressor:

Mean Squared Error: 0.1134

Mean Absolute Error: 0.2766

R^2 Score: 0.6106

Decision Tree Regressor:

Mean Squared Error: 0.1765

Mean Absolute Error: 0.3230

R^2 Score: 0.3940

K-Nearest Neighbors Regressor:

Mean Squared Error: 0.3704

Mean Absolute Error: 0.5185

R^2 Score: -0.2714

Support Vector Regressor:

Mean Squared Error: 0.2956

Mean Absolute Error: 0.4763

R^2 Score: -0.0145

Conclusion

Based on the evaluation metrics:

- Models like **Linear Regression** or **Gradient Boosting Regressor** may perform better for certain datasets due to their balance of complexity and accuracy.
- Ensemble methods like **Random Forest** and **Gradient Boosting** typically provide robust results with complex patterns.

- Simpler models like **K-Nearest Neighbors** or **Decision Tree Regressors** can be used for quick, interpretable solutions.