Model Validation, Overfitting Control & Hyperparameter Tuning

Submitted by: Aman Dwivedi

Problem & Overfitting Analysis

- **Objective**
- Improve model reliability using validation and tuning techniques.
- **Overfitting Detection**
- Unconstrained Decision Tree trained
- Train RMSE << Test RMSE
- Clear overfitting observed
- Conclusion: Model memorized training data and failed to generalize.

Cross-Validation

- Single train-test split is unreliable.
- Applied 5-Fold Cross Validation:
- Reduced bias from single split
- Obtained stable RMSE estimate
- Cross-validation confirmed model instability.

Hyperparameter Tuning

- Used **GridSearchCV** to tune:
- max_depth
- min_samples_split
- min_samples_leaf

Result:

- Reduced overfitting
- Improved generalization
- Balanced bias-variance tradeoff

Final Model Selection & Conclusion

- Compared:
- Linear Regression
- Ridge Regression
- Tuned Decision Tree
- Best Model: *Tuning decision tree*
- Reason:
- Lowest RMSE
- Highest R^2
- Stable cross-validation performance