**Report**

**Decision Tree Regressor Model Performance Analysis**

**For (startup.csv) dataset**

**Introduction**

The Decision Tree Regressor is a machine learning model that predicts values by repeatedly splitting the data into smaller parts based on certain rules. This report examines how different settings affect its accuracy to find the best setup.

**Methodology**

For the comparison of the performance of the DecisionTreeRegressor, we tried different settings for parameters:

* **Criterion:** The technique to compute the quality of a split ('squared error', 'friedman mse', 'absolute error', 'poisson').
* **Splitter:** Defines how the data is to be split ('best' or 'random').
* **Max Depth:** Defines the maximum depth of the tree.
* **Min Samples Split:** The minimum number of samples required to split an internal node.
* **Min Samples Leaf:** Minimum amount of samples to be at a leaf node.

**Other Parameters:**

* **Min Weight Fraction Leaf:** The minimum weight fraction of the total weights to be at a leaf node so as not to have nodes with few samples.
* **Max Features:** The number of features to employ when finding the best split ('None' employs all features, 'sqrt' employs square root of features, 'log2' employs logarithm of features).
* **Random State:** Controls randomness of the estimator, enabling reproducibility.
* **Max Leaf Nodes:** Limits the number of leaf nodes in the tree, controlling model complexity.
* **Min Impurity Decrease:** A split node threshold, where a node splits only if the impurity decrease is greater than this value.
* **CCP Alpha:** Pruning tree complexity parameter, determines the model-complexity vs. performance trade-off.
* **Monotonic Constraints:** Ensures monotonic feature relationships with the target variable, applies to certain domain constraints.

The default model was initially run followed by variations in every parameter to compare how each impacts performance.

**Results & Analysis**

Results for different parameter settings are described below:

**Default Model:** Without any parameter tuning, the model gave an R² score of 0.9291.

**Criterion:**

* 'squared error' gave 0.9366.
* 'friedman mse' gave 0.9420.
* 'absolute error' performed best with 0.9694.
* 'poisson' gave 0.9460.

**Splitter:**

* 'best' (default) gave 0.9361.
* 'random' gave 0.9174.

**Max Depth:**

* Default (None) gave 0.9326.

**Min Samples Split:**

* Default (2) gave 0.9220.

**Min Samples Leaf:**

* Default (1) gave 0.9248.

**Min Weight Fraction Leaf:**

* Default (0.0) gave 0.9301.

**Max Features:**

* Default (None) gave 0.9222.
* 'sqrt' obtained 0.7621.
* 'log2' obtained 0.8782.

**Random State:**

* Default (None) obtained 0.9398.

**Max Leaf Nodes:**

* Default (None) obtained 0.9240.

**Min Impurity Decrease:**

* Default (0.0) obtained 0.9455.

**CCP Alpha:**

* Default (0.0) obtained 0.9209.

**Monotonic Constraints:**

* Default (None) obtained 0.9211.

**Conclusion & Recommendations**

From the analysis:

* The **'absolute error'** criterion performed the best, with an R² score of 0. 0.9694.
* The 'poisson' criterion also performed well, with a score of 0.9460.
* The 'friedman mse' criterion provided a good balance, achieving 0.9420.
* The 'sqrt' option for max features resulted in a significant drop in performance (0.7621), while 'log2' performed moderately well (0.8782).
* Increasing min impurity decrease to 0.0 improved performance to 0.9455.
* Random state tuning led to a strong performance of 0.9398.

**Recommendations:**

For optimal performance, the DecisionTreeRegressor should use the 'absolute error' criterion and optimize parameters like random state, min impurity decrease, and max features carefully. Further tuning of parameters like CCP Alpha and monotonic constraints could also be explored for additional improvements.

This analysis provides valuable insights into parameter tuning for decision tree regression models, ensuring improved predictive performance.