**Predict Credit Consumption of Customer For a Leading Bank**

Model Documentation:

The model used for this analysis is a Random Forest regressor. Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It is widely used for regression and classification tasks due to its effectiveness and ability to handle complex data.

Model Architecture:

The Random Forest model consists of multiple decision trees, where each tree is trained on a random subset of the training data with replacement (bootstrap samples) and a random subset of features. The predictions from all the trees are then combined through averaging (regression) or voting (classification) to obtain the final prediction.

Hyperparameters:

The following hyperparameters were used for the Random Forest model:

**n\_estimators:** The number of decision trees in the forest. In this case, it was set to 100, indicating that 100 trees were used in the ensemble.

**max\_depth**: The maximum depth of each decision tree. Setting it to None allows the tree to grow until all the leaves are pure or until all leaves contain fewer samples than **min\_samples\_split.** This enables the trees to capture complex relationships in the data.

**min\_samples\_split:** The minimum number of samples required to split an internal node. It was set to 2, meaning that a node will be split if it contains at least 2 samples.

**min\_samples\_leaf:** The minimum number of samples required to be at a leaf node. It was set to 1, indicating that even the leaves with a single sample will be considered.

Model Evaluation:

The model's performance was evaluated using cross-validation. The dataset was divided into multiple subsets or folds, and the model was trained and evaluated on each fold. The evaluation metric used was the R-squared score, which measures the proportion of the variance in the target variable that can be explained by the model. A higher R-squared score indicates a better fit of the model to the data.

Insights and Observations**:**

The Random Forest model was able to capture complex relationships in the data and provide accurate predictions. The mean R-squared score obtained through cross-validation provides an estimate of the model's generalization performance. The standard deviation of the R-squared scores reflects the variability in the model's performance across different folds, indicating the stability of the model.

By using cross-validation, we can have a more reliable assessment of the model's performance, as it accounts for the potential variation in the data and reduces the risk of overfitting to a specific training-test split.

Overall, the Random Forest model with the chosen hyperparameters and cross-validation approach is effective in predicting the target variable and provides a reliable estimate of its performance on unseen data.