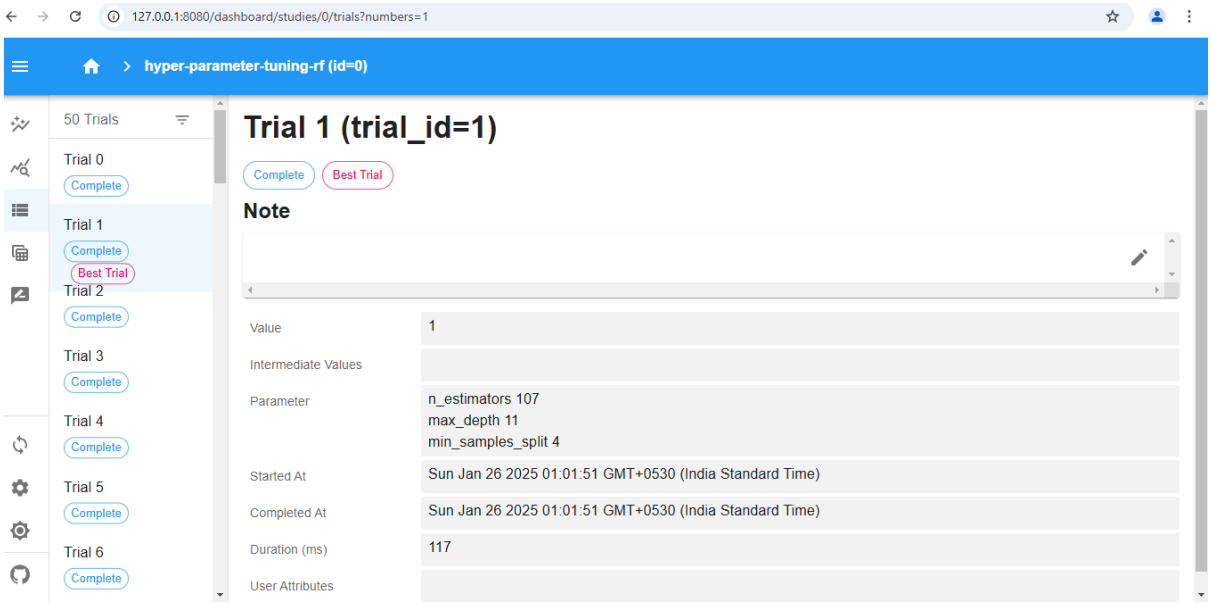
**Hyper-Parameter Tuning Report.**

For this assignment, we have chosen the first 3 hyper-parameters namely

- **n_estimators.**
- **max_depth**
- **min_samples_split**

The best parameter was found in the second attempt (attempt 1, after attempt 0) with an accuracy of 1. The hyper-parameter values obtained are: **'n_estimators': 66, 'max_depth': 8, 'min_samples_split': 3**.

Here is an **Obtune-Dashboard Graph** showing the comparisons and the **Trial 1** as the Best Trial.




**Logs Collected:**

[I 2025-01-26 01:12:10,712] A new study created in Journal with name: hyper-parameter-tuning-rf

[I 2025-01-26 01:12:10,744] Trial 0 finished with value: 0.9666666666666667 and parameters: {'n_estimators': 3, 'max_depth': 4, 'min_samples_split': 4}. Best is trial 0 with value: 0.9666666666666667.

[I 2025-01-26 01:12:10,839] Trial 1 finished with value: 1.0 and parameters: {'n_estimators': 66, 'max_depth': 8, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:10,886] Trial 2 finished with value: 1.0 and parameters: {'n_estimators': 23, 'max_depth': 4, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:10,982] Trial 3 finished with value: 1.0 and parameters: {'n_estimators': 85, 'max_depth': 13, 'min_samples_split': 2}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,097] Trial 4 finished with value: 1.0 and parameters: {'n_estimators': 90, 'max_depth': 9, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,173] Trial 5 finished with value: 1.0 and parameters: {'n_estimators': 41, 'max_depth': 4, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,252] Trial 6 finished with value: 1.0 and parameters: {'n_estimators': 55, 'max_depth': 8, 'min_samples_split': 2}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,316] Trial 7 finished with value: 1.0 and parameters: {'n_estimators': 48, 'max_depth': 5, 'min_samples_split': 5}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,364] Trial 8 finished with value: 1.0 and parameters: {'n_estimators': 18, 'max_depth': 4, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,476] Trial 9 finished with value: 1.0 and parameters: {'n_estimators': 90, 'max_depth': 5, 'min_samples_split': 5}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,571] Trial 10 finished with value: 1.0 and parameters: {'n_estimators': 67, 'max_depth': 13, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,634] Trial 11 finished with value: 1.0 and parameters: {'n_estimators': 31, 'max_depth': 7, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,746] Trial 12 finished with value: 1.0 and parameters: {'n_estimators': 68, 'max_depth': 10, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,809] Trial 13 finished with value: 1.0 and parameters: {'n_estimators': 23, 'max_depth': 11, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,917] Trial 14 finished with value: 1.0 and parameters: {'n_estimators': 69, 'max_depth': 7, 'min_samples_split': 2}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:11,996] Trial 15 finished with value: 1.0 and parameters: {'n_estimators': 35, 'max_depth': 15, 'min_samples_split': 5}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:12,044] Trial 16 finished with value: 1.0 and parameters: {'n_estimators': 3, 'max_depth': 6, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:12,155] Trial 17 finished with value: 1.0 and parameters: {'n_estimators': 59, 'max_depth': 3, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:12,266] Trial 18 finished with value: 1.0 and parameters: {'n_estimators': 76, 'max_depth': 11, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:12,330] Trial 19 finished with value: 1.0 and parameters: {'n_estimators': 17, 'max_depth': 8, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:12,410] Trial 20 finished with value: 1.0 and parameters: {'n_estimators': 45, 'max_depth': 3, 'min_samples_split': 2}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:12,538] Trial 21 finished with value: 1.0 and parameters: {'n_estimators': 97, 'max_depth': 13, 'min_samples_split': 2}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:12,674] Trial 22 finished with value: 1.0 and parameters: {'n_estimators': 78, 'max_depth': 15, 'min_samples_split': 2}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:12,792] Trial 23 finished with value: 1.0 and parameters: {'n_estimators': 83, 'max_depth': 13, 'min_samples_split': 2}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:12,887] Trial 24 finished with value: 1.0 and parameters: {'n_estimators': 62, 'max_depth': 11, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,015] Trial 25 finished with value: 1.0 and parameters: {'n_estimators': 99, 'max_depth': 9, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,141] Trial 26 finished with value: 1.0 and parameters: {'n_estimators': 85, 'max_depth': 12, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,269] Trial 27 finished with value: 1.0 and parameters: {'n_estimators': 75, 'max_depth': 6, 'min_samples_split': 2}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,362] Trial 28 finished with value: 1.0 and parameters: {'n_estimators': 57, 'max_depth': 14, 'min_samples_split': 5}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,411] Trial 29 finished with value: 1.0 and parameters: {'n_estimators': 5, 'max_depth': 10, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,489] Trial 30 finished with value: 1.0 and parameters: {'n_estimators': 37, 'max_depth': 8, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,614] Trial 31 finished with value: 1.0 and parameters: {'n_estimators': 93, 'max_depth': 9, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,731] Trial 32 finished with value: 1.0 and parameters: {'n_estimators': 86, 'max_depth': 10, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,839] Trial 33 finished with value: 1.0 and parameters: {'n_estimators': 78, 'max_depth': 7, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:13,945] Trial 34 finished with value: 1.0 and parameters: {'n_estimators': 53, 'max_depth': 8, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,059] Trial 35 finished with value: 1.0 and parameters: {'n_estimators': 88, 'max_depth': 5, 'min_samples_split': 5}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,202] Trial 36 finished with value: 1.0 and parameters: {'n_estimators': 93, 'max_depth': 4, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,298] Trial 37 finished with value: 1.0 and parameters: {'n_estimators': 65, 'max_depth': 12, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,377] Trial 38 finished with value: 1.0 and parameters: {'n_estimators': 49, 'max_depth': 6, 'min_samples_split': 5}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,488] Trial 39 finished with value: 1.0 and parameters: {'n_estimators': 73, 'max_depth': 9, 'min_samples_split': 2}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,551] Trial 40 finished with value: 1.0 and parameters: {'n_estimators': 10, 'max_depth': 4, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,614] Trial 41 finished with value: 1.0 and parameters: {'n_estimators': 26, 'max_depth': 5, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,677] Trial 42 finished with value: 1.0 and parameters: {'n_estimators': 39, 'max_depth': 3, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,772] Trial 43 finished with value: 1.0 and parameters: {'n_estimators': 43, 'max_depth': 4, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,835] Trial 44 finished with value: 1.0 and parameters: {'n_estimators': 29, 'max_depth': 5, 'min_samples_split': 5}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:14,899] Trial 45 finished with value: 1.0 and parameters: {'n_estimators': 18, 'max_depth': 7, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:15,010] Trial 46 finished with value: 1.0 and parameters: {'n_estimators': 70, 'max_depth': 6, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:15,121] Trial 47 finished with value: 1.0 and parameters: {'n_estimators': 83, 'max_depth': 10, 'min_samples_split': 3}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:15,201] Trial 48 finished with value: 1.0 and parameters: {'n_estimators': 33, 'max_depth': 3, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

[I 2025-01-26 01:12:15,328] Trial 49 finished with value: 1.0 and parameters: {'n_estimators': 92, 'max_depth': 11, 'min_samples_split': 4}. Best is trial 1 with value: 1.0.

Best hyperparameters: {'n_estimators': 66, 'max_depth': 8, 'min_samples_split': 3}

Best accuracy: 1.0

**Hyper-Parameters for Random Forest Classifier.**

For a ***RandomForestClassifier*** Model, the hyper-parameters are:

- **n_estimators:** Random forest is a group of many decision trees, the n_estimator parameter controls the number of trees inside the classifier. Ideally using many trees to fit a model will help us to get a more generalized result, but this is not always the case. However, it will not cause any overfitting but can certainly increase the time complexity of the model. **The default number of estimators is 100** in scikit-learn.
- **max_depth:** It governs the maximum height upto which the trees inside the forest can grow. It is one of the most important hyperparameters when it comes to increasing the accuracy of the model, as we increase the depth of the tree the model accuracy increases upto a certain limit but then it will start to decrease gradually because of overfitting in the model. It is important to set its value appropriately to avoid overfitting. **The default value is set to None,** None specifies that the nodes inside the tree will continue to grow until all leaves become pure or all leaves contain less than **min_samples_split** (another hyperparameter).
- **min_samples_split:** It specifies the minimum amount of samples an internal node must hold in order to split into further nodes. If we have a very low value of min_samples_splits then, in this case, our tree will continue to grow and start overfitting. By increasing the value of min_samples_splits we can decrease the total number of splits thus limiting the number of parameters in the model and thus can aid in reducing the overfitting in the model. However, the value should not be kept very large that a number of parameters drop extremely causing the model to underfit. We generally keep min_samples_split value between 2 and 6. **However, the default value is set to 2.**

- **min_samples_leaf:** It specifies the minimum amount of samples that a node must hold after getting split. It also helps to reduce overfitting when we have ample amount of parameters. Less number of parameters can lead to overfitting also, we should keep in mind that increasing the value to a large number can lead to less number of parameters and in this case model can underfit also. **The default value is set to 1.**
- **max_features:** Random forest takes random subsets of features and tries to find the best split. max_features helps to find the number of features to take into account in order to make the best split. It can take four values "**auto**", "**sqrt**", "**log2**" and **None**.
    - In case of auto: considers max_features = sqrt(n_features)
    - In case of sqrt: considers max_features = sqrt(n_features), it is same as auto
    - In case of log2: considers max_features = log2(n_features)
    - In case of None: considers max_features = n_features
- **max_leaf_nodes:** It sets a limit on the splitting of the node and thus helps to reduce the depth of the tree, and effectively helps in reducing overfitting. If the value is set to None, the tree continues to grow infinitely.
- **max_samples:** This hyperparameter helps to choose maximum number of samples from the training dataset to train each individual tree.