

Create a Simple Classifier (Random Forest)

- Launch Rapid Miner
- File → New Process
- Load a CSV File (PlayerPerformance)

Create Random Forest(RF) Classifier

- Write *Random Forest* in search box (Modeling → Classification & Regression → Tree Induction → Random Forest)
Drag *Random Forest* operator and drop it to the main process area

Select RF operator and Tune the Parameters → click Play

The screenshot displays the RapidMiner software interface. On the left, a workflow is visible in the 'Process' pane, consisting of a 'Read CSV' operator connected to a 'Random Forest' operator. The 'Random Forest' operator is highlighted with an orange border. On the right, the 'Parameters' pane is open, showing the configuration for the 'Random Forest' operator. The parameters are as follows:

Parameter	Value
number of trees	10
criterion	gain_ratio
maximal depth	20
apply pruning	<input checked="" type="checkbox"/>
confidence	0.25
apply prepruning	<input checked="" type="checkbox"/>
minimal gain	0.1

At the bottom of the parameters pane, there is a 'Help' button and a 'Play' button (represented by a green triangle). The title bar of the parameters pane reads 'Random Forest (RapidMiner Studio GUI)'.

Random Forest (Multiple Trees) is built.

The screenshot displays a software interface for a Random Forest model. The top bar shows two tabs: "Result Overview" and "Random Forest Model (Random Forest)". The left sidebar lists ten models, all labeled "Model 1 (Random Forest)" through "Model 10 (Random Forest)". The main area is divided into three sections: "Graph", "Description", and "Annotation". The "Graph" section contains a "Zoom" control with magnifying glass icons and a "Mode" section with a "Tree" dropdown menu. The "Description" section has a "Save Image..." button. The "Annotation" section has a "Help" button. The "Tree" dropdown is expanded, showing a decision tree structure. The tree starts with a root node "Outlook", which branches into three nodes: "= overc", "= rainy", and "= sunny". The "= overc" and "= rainy" nodes lead to leaf nodes labeled "P", each with a red bar. The "= sunny" node leads to a "Temperature" node, which branches into two nodes: "= |=" and "moderate". The "= |=" node leads to a leaf node labeled "N" with a blue bar, and the "moderate" node leads to a leaf node labeled "P" with a red bar.

Random Forest Model (Random Forest)

Model 1 (Random Forest)

Model 2 (Random Forest)

Model 3 (Random Forest)

Model 4 (Random Forest)

Model 5 (Random Forest)

Model 6 (Random Forest)

Model 7 (Random Forest)

Model 8 (Random Forest)

Model 9 (Random Forest)

Model 10 (Random Forest)

Graph

Zoom

Mode

Tree

Node Labels

Edge Labels

Save Image...

Help

Outlook

= overc = rainy = sunny

P

P

Temperature

= | = moderate

N

P

This is the simplest form of RF modeling.

This RF uses all rows of training data set for learning.

Objectives

- Explain RF and its hyper parameters. (in single statement).
- What are the two most important hyper parameters of RF? Justify. (Please don't google, I need your opinion)
- Make RF classifier for your data set.
- Explain a tree from RF. Justify the selection criteria.
- Extract Good Rules from the selected tree.
- Apply prediction on five random selected rows.
- Calculate the Prediction Accuracy.

Due on 13th March (till 12) *Good Luck*