# Pre-processing Techniques Applied

### Filter Examples

* Used to filter out rows with missing data.

### Select Attribute

* Used to select only the required attributes for the model training.

### Set Role

* Sets the role of the output attribute as the label

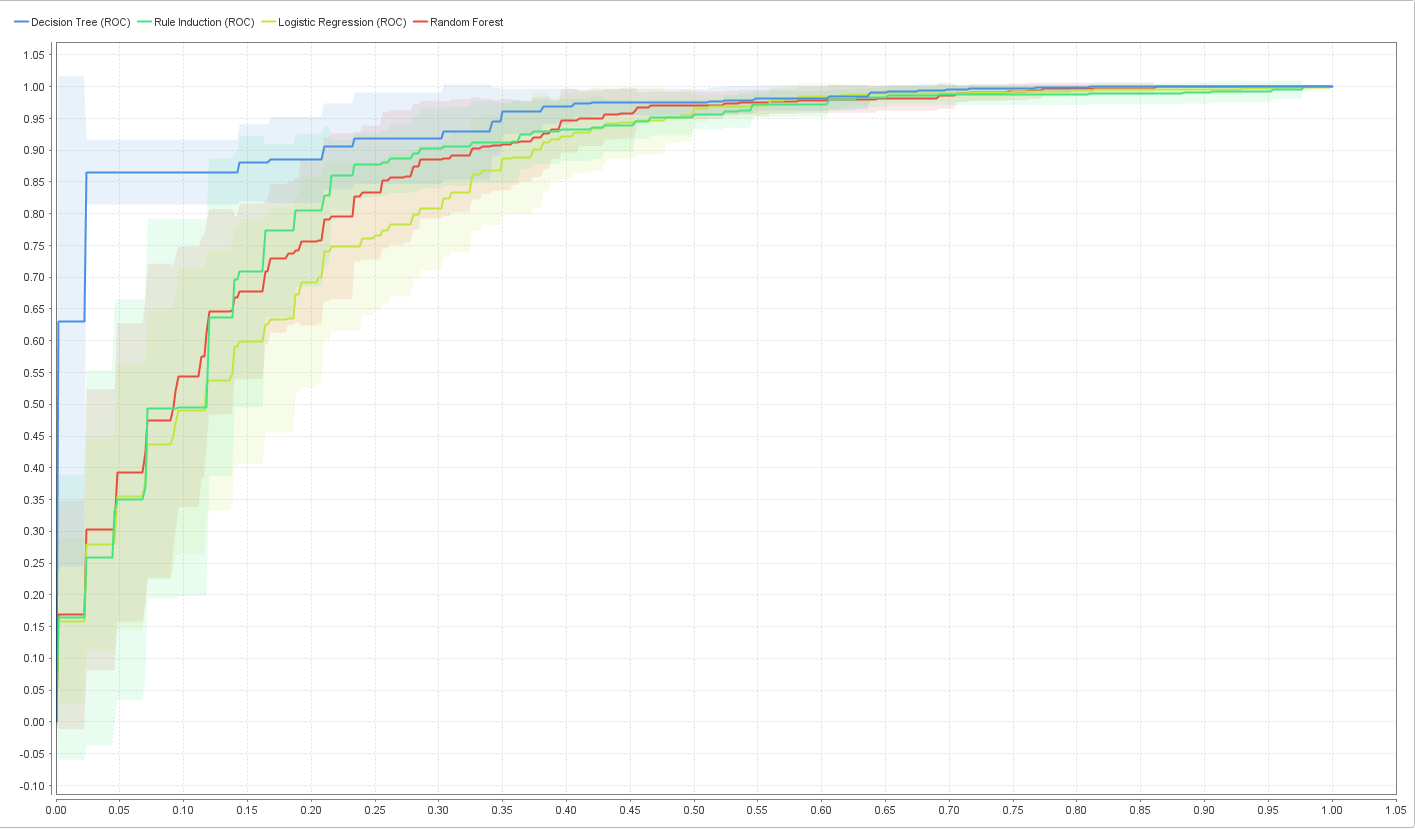
### Multiply

* Duplicates the original data to be used in model training, cross validation and compare ROCs operators.

### Split Data

* Splits the data such that all the models are trained with the same data for proper accuracy measurement in the output.

# ROC Curve



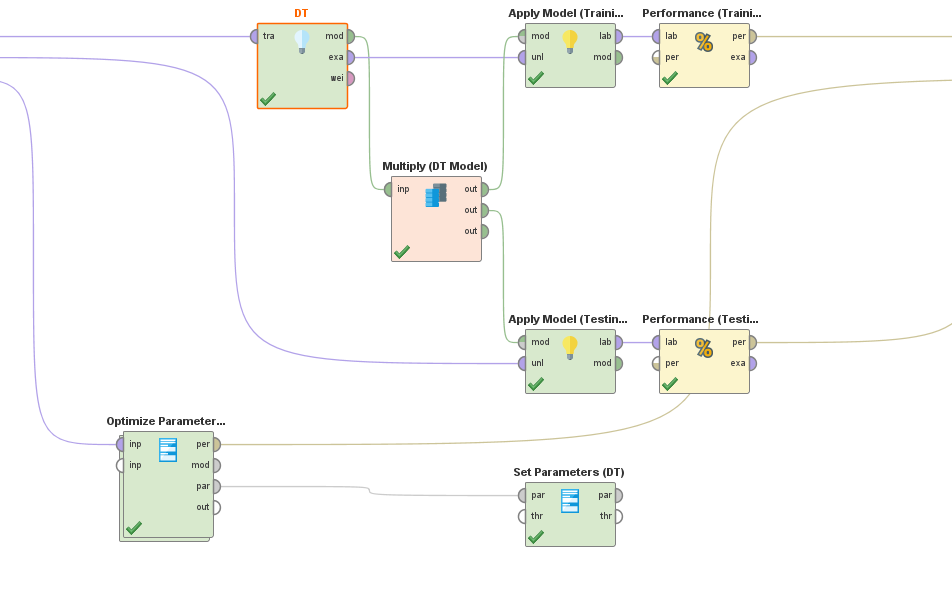
*(fig 1: ROC curve)*

# Technique 1: Decision Tree

### Motivation

* Dataset has binomial output/ class i.e., discrete data. So, classification algorithms are best for the dataset and decision tree is a classification technique
* Best Area under the ROC Curve (fig 1)

### Snapshot

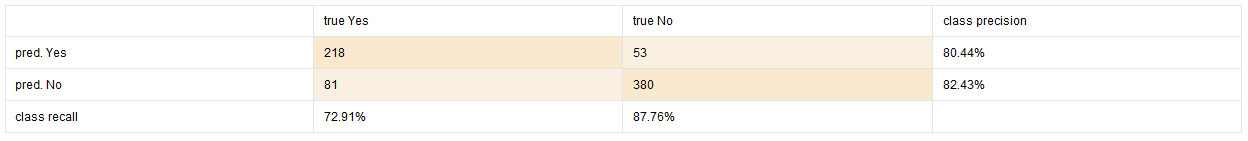


### Parameter Settings

* Criterion: gini\_index
* Maximal depth: 4
* Apply pruning: false
* Apply prepruning: false

**Optimize Parameters** operator used for the best performance.

|  |  |
| --- | --- |
| accuracy | 81.69% |
| precision | 82.43% (positive class: No) |
| recall | 87.76% (positive class: No) |
| f-measure | 85.01% (positive class: No) |

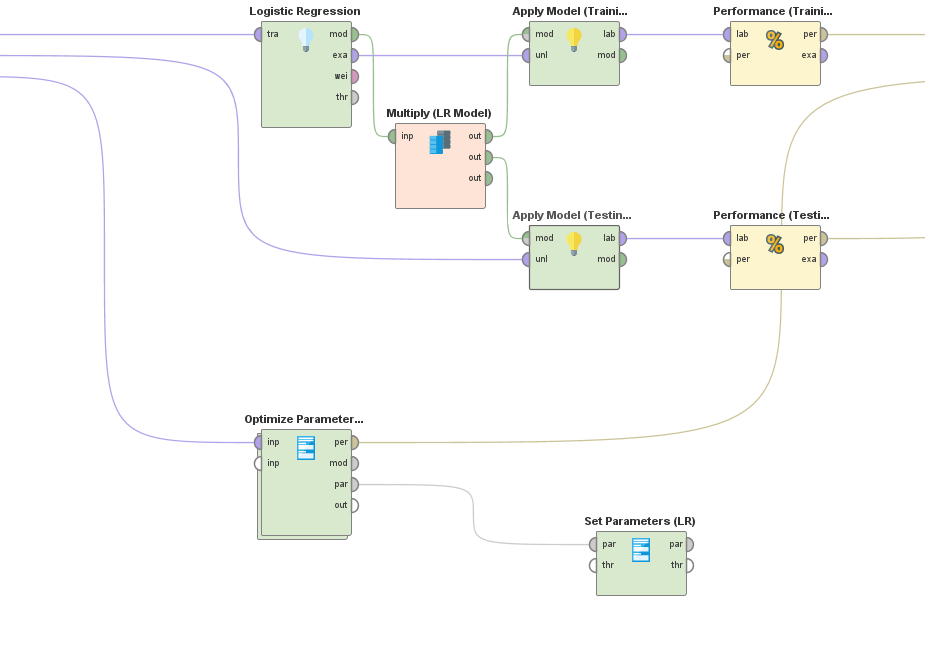


# Technique 2: Logistic Regression

### Motivation

* Dataset has binomial output/ class i.e., discrete data. So, classification algorithms are best for the dataset and Logistic Regression is a classification technique
* Complex algorithm/ model (requirement)

### Snapshot

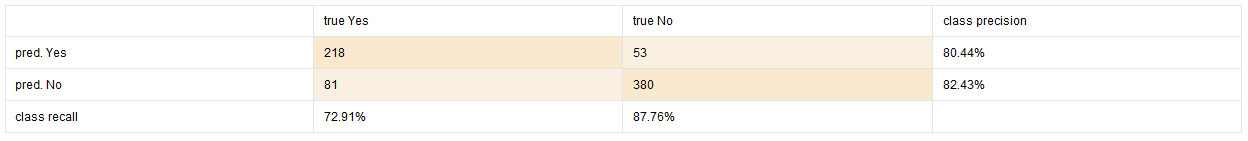


### Parameter Settings

* Solver: auto
* Reproducible: false
* User regularization: false
* Standardize: true
* Non-negative coefficients: false
* Add intercept: true
* Compute p-values: true
* Remove collinear column: true

**Optimize Parameters** operator used for the best result.

|  |  |
| --- | --- |
| accuracy | 78.01% |
| precision | 80.09% (positive class: No) |
| recall | 83.60% (positive class: No) |
| f-measure | 81.81% (positive class: No) |



# Comparison of Testing Performance

### Testing Performance Report

### Cross-validation Performance Report