

## Exercise

A standard classification evaluation process consists of several steps

1. Build a classifier
2. Evaluate a classifier
3. Compare performance of several classifiers using charts such as ROC, Precision vs. Recall, Lift Charts

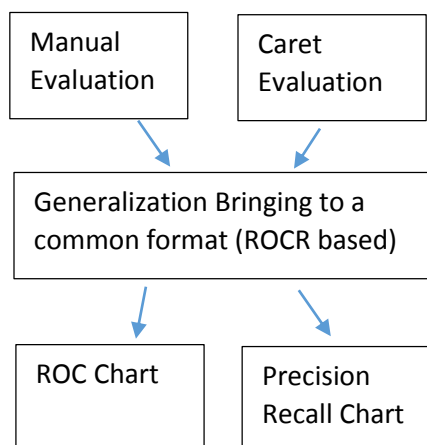
The task is to simplify the process by providing a generic approach for comparing classifier performances in R.

Why the generic approach is required? R has many algorithms that can be run manually. R has also packages that wrap the complexity of running each classification algorithm manually. The output of running the classification algorithm manually or running a wrapper package is different but we would like to have the flexibility to use both approaches and quickly show results (save our time by repeating the same process again and again)

The main wrapper package in R is **caret**. The algorithms we would like to test in this exercise are Logistic Regression (**glm**) and Random Forest (**randomForest**) but in reality we can test any classification algorithm.

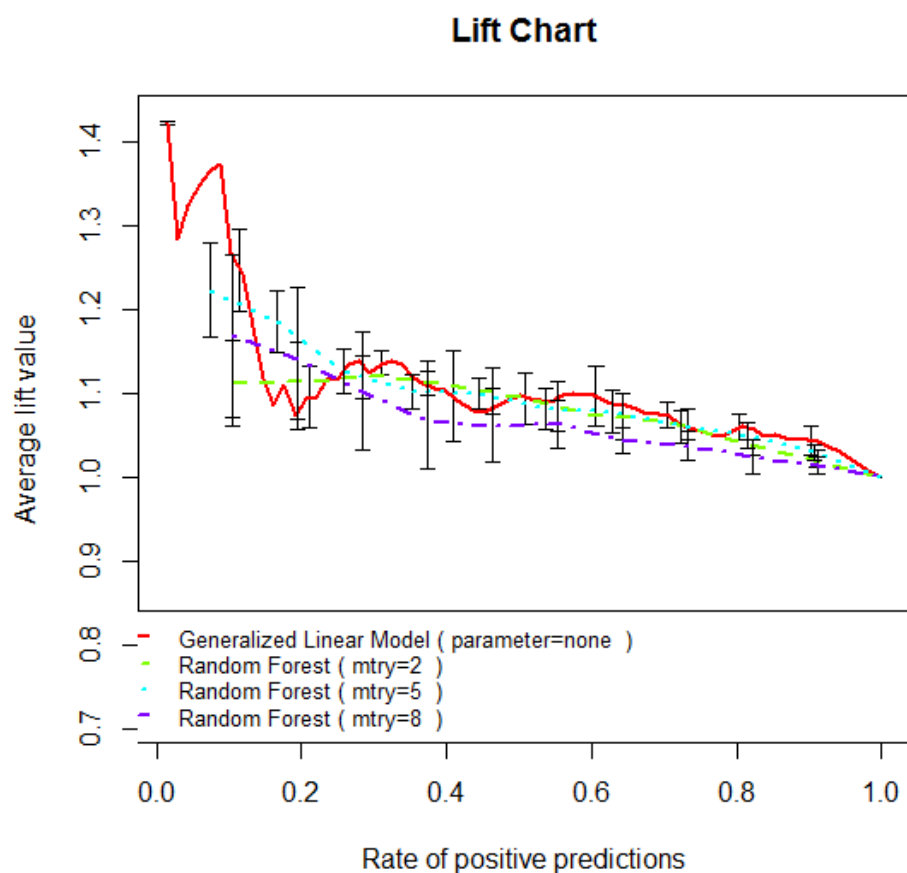
The package for Visual Classifier Evaluation is **ROCR**

The schematic diagram of the task is shown below



Output example: In this example we evaluated two algorithms – glm and Random Forest using 5 fold cross validation (therefore one can see SD) . Caret by default is using several mtry parameters for random forest – in this case three 5 fold cross validations were performed by caret and for each mtry there is a lift chart.

Logistic regression doesn't have any control parameters therefore parameter = none. This information is stored in the caret object.



Requirements:

1. Write a function that converts the results of manual classifier into the generic representation
2. Write a function that converts the results of caret-based classification into the generic representation
3. Use the generic representation to plot graphs for ROC, Precision-Recall, Lift
4. The number of algorithms to visualize is not fixed (can be one or more, so it should be flexible – use list object for this)

Notes: **dplyr** package is very useful for preparing the generic representation. Packages needed: caret, ROCR, randomForest