

## **Ensemble Modeling Workshop**

It is a continuation of the previous modelling workshop using the same dataset – TelcoChurn. Follow the instructions below carefully to build RandomForest and Boosting models.

### **Instructions:**

1. Start RStudio and Rattle.  
    > library(rattle)  
    > rattle()
2. Locate the TelcoChurn dataset: TelcoChurn.csv
3. Partition the dataset as default: 70/15/15, meaning 70% for training, 15% for validation, and 15% for testing. (or try 50/0/50)
4. Click the Execute button to load the data.
5. Check out the different roles assigned to Variables: input, target, or identification.
6. Click the Model tab; Select FOREST model; Select Traditional algorithm: randomForest();  
    Check out the default settings:  
    Number of Trees, Number of variables, etc.
7. Click the Execute button to build a Random Forest model.
8. Browse the Modeling results.
9. Click the Importance, Errors, Rules, OOB ROC buttons to explore the results further.
10. Experiment with different parameter settings.
11. Click the Model tab; Select BOOST model;  
    Select **Adaptive** algorithm: ada()  
    Check out the default settings:  
    Number of Trees, Max Depth, Min Split, Complexity, etc.
12. Click the Execute button to build a Boosting model.
13. Browse the Modeling results

14. Click the Importance and Errors buttons to explore the results further.
15. Experiment with different parameter settings.
16. Click the Evaluate tab.
17. Select FOREST and BOOST models; select different evaluation types such as Error Matrix (Confusion Matrix), ROC, etc.
18. Identify the dataset for evaluation, usually the testing dataset.
19. Click the Execute button to view the evaluation results.
20. Compare the performance of these models, may also include other models such as decision tree, neural networks and SVM.
21. Check out the R code generated in the Log tab.
22. Save your project.