Project 4: Human Resources

**Problem Statement**

Why are our best and most experienced employees leaving prematurely? Use this database and try to predict which valuable employees will leave next. Fields in the dataset include:

* Employee satisfaction level
* Last evaluation
* Number of projects
* Average monthly hours
* Time spent at the company
* Whether they have had a work accident
* Whether they have had a promotion in the last 5 years
* Department
* Salary
* Whether the employee has left

We are given two datasets , hr\_train.csv and hr\_test.csv . We need to use data hr\_train to build predictive model for response variable ‘left’. hr\_test data contains all other factors except “left”, we need to predict that using the model that we developed and submit the predicted values in a csv files.

We have to submit the probability scores, not the hard classes.

If we are using decision trees or random forest here, probability scores can be calculated as:

score=predict(rf\_model,newdata= testdata, type="prob")[,1]

score=predict(tree\_model,newdata= testdata, type=‘vector’)[,1]

Evaluation Criterion : auc score on test data. larger auc score, better Model

1. Our auc score for test data should come out to be more than 0.84

2. Your predictions should not contain any NA values.