MNIST with H2O and RandomForest

#Learnt h2o usage from the below link #https://www.analyticsvidhya.com/blog/2016/05/h2o-data-table-build-models-large-da

##RandomForest with untouched raw pixel image

```
library(h2o)
```

```
##
## ----------------------------------------------------------------------
##
## Your next step is to start H2O:
##     > h2o.init()
##
## For H2O package documentation, ask for help:
##     > ??h2o
##
## After starting H2O, you can use the Web UI at http://localhost:54321
## For more information visit http://docs.h2o.ai
##
## ----------------------------------------------------------------------
```

```
##
## Attaching package: 'h2o'
```

```
## The following objects are masked from 'package:stats':
##
##     cor, sd, var
```

```
## The following objects are masked from 'package:base':
##
##     &&, %*%, %in%, ||, apply, as.factor, as.numeric, colnames,
##     colnames<-, ifelse, is.character, is.factor, is.numeric, log,
##     log10, log1p, log2, round, signif, trunc
```

```
train = read.csv("mnist_train.csv", header = TRUE)
test = read.csv("mnist_test.csv", header = TRUE)
h2o.init(nthreads = -1, max_mem_size = '4g', ip = "127.0.0.1", port = 50001)
```

```
##
## H2O is not running yet, starting it now...
##
## Note:  In case of errors look at the following log files:
##     /var/folders/7d/fzx0z4t54zj78z__kpp0vzh00000gn/T//RtmpgqJp1u/h2o_amishukl_started_from_r.out
##     /var/folders/7d/fzx0z4t54zj78z__kpp0vzh00000gn/T//RtmpgqJp1u/h2o_amishukl_started_from_r.err
##
##
## Starting H2O JVM and connecting: .. Connection successful!
##
## R is connected to the H2O cluster:
##     H2O cluster uptime:         2 seconds 4 milliseconds
##     H2O cluster timezone:       America/Los_Angeles
##     H2O data parsing timezone:  UTC
##     H2O cluster version:        3.26.0.2
##     H2O cluster version age:    2 months and 7 days
##     H2O cluster name:           H2O_started_from_R_amishukl_scx813
```

```
##     H2O cluster total nodes:    1
##     H2O cluster total memory:   4.00 GB
##     H2O cluster total cores:    8
##     H2O cluster allowed cores:  8
##     H2O cluster healthy:        TRUE
##     H2O Connection ip:          127.0.0.1
##     H2O Connection port:        50001
##     H2O Connection proxy:       NA
##     H2O Internal Security:      FALSE
##     H2O API Extensions:         Amazon S3, XGBoost, Algos, AutoML, Core V3, Core V4
##     R Version:                  R version 3.5.2 (2018-12-20)
```

```
train$class = as.factor(train$class)
test$class = as.factor(test$class)

train.h2o <- as.h2o(train)
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
test.h2o <- as.h2o(test)
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
T.dep <- 785
T.indep <- c(1:784)

#Tree 10, depth 4
model <- h2o.randomForest(y=T.dep, x=T.indep, training_frame = train.h2o,
                          ntrees = 10, max_depth = 4, seed = 1234)
```

```
## Warning in .h2o.startModelJob(algo, params, h2oRestApiVersion): Dropping bad and constant columns: [
```

```
##
  |
  |                                                                      |   0%
  |
  |======                                                                |  10%
  |
  |=============                                                         |  20%
  |
  |=========================                                             |  40%
  |
  |==============================                                        |  50%
  |
  |=============================================                         |  70%
  |
  |=================================================================     |  90%
  |
  |======================================================================| 100%
```

```
preds <- as.data.frame(h2o.predict(model, newdata=test.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(test$class, as.vector(preds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.8585
```

```
#Tree 10, depth 8
model <- h2o.randomForest(y=T.dep, x=T.indep, training_frame = train.h2o, ntrees = 10, max_depth = 8, s
```

```
## Warning in .h2o.startModelJob(algo, params, h2oRestApiVersion): Dropping bad and constant columns: [
```

```
##
  |
  |                                                                      |   0%
  |
  |======                                                                |  10%
  |
  |=============                                                         |  20%
  |
  |====================                                                  |  30%
  |
  |===========================                                           |  40%
  |
  |==================================                                    |  50%
  |
  |==========================================                            |  60%
  |
  |===================================================                   |  70%
  |
  |==========================================================            |  80%
  |
  |======================================================================| 100%
```

```
preds <- as.data.frame(h2o.predict(model, newdata=test.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(test$class, as.vector(preds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9311
```

```
#Tree 10, depth 16
model <- h2o.randomForest(y=T.dep, x=T.indep, training_frame = train.h2o,
                          ntrees = 10, max_depth = 16, seed = 1234)
```

```
## Warning in .h2o.startModelJob(algo, params, h2oRestApiVersion): Dropping bad and constant columns: [p

##
  |
  |                                                                      |   0%
  |
  |======                                                                |  10%
  |
  |=============                                                         |  20%
  |
  |====================                                                  |  30%
  |
  |===========================                                           |  40%
  |
  |==================================                                    |  50%
  |
  |==========================================                            |  60%
  |
  |=================================================                     |  70%
  |
  |========================================================              |  80%
  |
  |===============================================================       |  90%
  |
  |======================================================================| 100%
```

```
preds <- as.data.frame(h2o.predict(model, newdata=test.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(test$class, as.vector(preds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9559
```

```
#Tree 20, depth 4
model <- h2o.randomForest(y=T.dep, x=T.indep, training_frame = train.h2o,
                          ntrees = 20, max_depth = 4, seed = 1234)
```

```
## Warning in .h2o.startModelJob(algo, params, h2oRestApiVersion): Dropping bad and constant columns: [p

##
  |
  |                                                                      |   0%
  |
  |==========                                                            |  15%
  |
  |====================                                                  |  30%
  |
  |==============================                                        |  45%
  |
```

```
  |=====================================                                 |  60%
  |
  |================================================                      |  75%
  |
  |============================================================          |  90%
  |
  |======================================================================| 100%
```

```
preds <- as.data.frame(h2o.predict(model, newdata=test.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(test$class, as.vector(preds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.8715
```

```
#Tree 20, depth 8
model <- h2o.randomForest(y=T.dep, x=T.indep, training_frame = train.h2o,
                          ntrees = 20, max_depth = 8, seed = 1234)
```

```
## Warning in .h2o.startModelJob(algo, params, h2oRestApiVersion): Dropping bad and constant columns: [
```

```
##
  |
  |                                                                      |   0%
  |
  |===                                                                   |   5%
  |
  |======                                                                |  10%
  |
  |==========                                                            |  15%
  |
  |=============                                                         |  20%
  |
  |====================                                                  |  30%
  |
  |========================                                              |  35%
  |
  |============================                                          |  40%
  |
  |================================                                      |  45%
  |
  |===================================                                   |  50%
  |
  |=======================================                               |  60%
  |
  |==============================================                        |  65%
  |
  |==================================================                    |  70%
  |
  |======================================================                |  75%
```

```
  |
  |=================================================                     |  80%
  |
  |=====================================================                 |  85%
  |
  |=============================================================         |  95%
  |
  |=====================================================================| 100%
```

```
preds <- as.data.frame(h2o.predict(model, newdata=test.h2o))
```

```
##
  |
  |                                                                     |   0%
  |
  |=====================================================================| 100%
```

```
confusionMatrix <- table(test$class, as.vector(preds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9386
```

```
#Tree 20, depth 16
model <- h2o.randomForest(y=T.dep, x=T.indep, training_frame = train.h2o,
                          ntrees = 20, max_depth = 16, seed = 1234)
```

```
## Warning in .h2o.startModelJob(algo, params, h2oRestApiVersion): Dropping bad and constant columns: [p
```

```
##
  |
  |                                                                     |   0%
  |
  |===                                                                  |   5%
  |
  |======                                                               |  10%
  |
  |==========                                                           |  15%
  |
  |=============                                                        |  20%
  |
  |================                                                     |  25%
  |
  |====================                                                 |  30%
  |
  |=======================                                              |  35%
  |
  |===========================                                          |  40%
  |
  |===============================                                      |  45%
  |
  |==================================                                   |  50%
  |
  |======================================                               |  55%
  |
  |==========================================                           |  60%
  |
```

```
  |=======================================                          |  65%
  |
  |=========================================                        |  70%
  |
  |===========================================                      |  75%
  |
  |=============================================                    |  80%
  |
  |===============================================                  |  85%
  |
  |=================================================                |  90%
  |
  |===================================================              |  95%
  |
  |=================================================================| 100%
```

```
preds <- as.data.frame(h2o.predict(model, newdata=test.h2o))
```

```
##
  |
  |                                                                 |   0%
  |
  |=================================================================| 100%
```

```
confusionMatrix <- table(test$class, as.vector(preds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9641
```

```
#Tree 30, depth 4
model <- h2o.randomForest(y=T.dep, x=T.indep, training_frame = train.h2o,
                          ntrees = 30, max_depth = 4, seed = 1234)
```

```
## Warning in .h2o.startModelJob(algo, params, h2oRestApiVersion): Dropping bad and constant columns: [
```

```
##
  |
  |                                                                 |   0%
  |
  |====                                                             |   7%
  |
  |=========                                                        |  13%
  |
  |=============                                                    |  20%
  |
  |================                                                 |  23%
  |
  |====================                                             |  30%
  |
  |========================                                         |  37%
  |
  |============================                                     |  43%
  |
  |================================                                 |  50%
  |
  |====================================                             |  57%
```

```
  |
  |=======================================                          |  63%
  |
  |============================================                     |  70%
  |
  |================================================                 |  77%
  |
  |====================================================             |  80%
  |
  |========================================================         |  87%
  |
  |============================================================     |  93%
  |
  |================================================================| 100%
```

```
preds <- as.data.frame(h2o.predict(model, newdata=test.h2o))
```

```
##
  |
  |                                                                 |   0%
  |
  |================================================================| 100%
```

```
confusionMatrix <- table(test$class, as.vector(preds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.8771
```

```
#Tree 30, depth 8
model <- h2o.randomForest(y=T.dep, x=T.indep, training_frame = train.h2o,
                          ntrees = 30, max_depth = 8, seed = 1234)
```

```
## Warning in .h2o.startModelJob(algo, params, h2oRestApiVersion): Dropping bad and constant columns: [
```

```
##
  |
  |                                                                 |   0%
  |
  |==                                                               |   3%
  |
  |====                                                             |   7%
  |
  |======                                                           |  10%
  |
  |=========                                                        |  13%
  |
  |===========                                                      |  17%
  |
  |=============                                                    |  20%
  |
  |===============                                                  |  23%
  |
  |=================                                                |  27%
  |
  |====================                                             |  30%
  |
```

```
  |======================                                                |  33%
  |
  |========================                                              |  37%
  |
  |==========================                                            |  40%
  |
  |============================                                          |  43%
  |
  |==============================                                        |  47%
  |
  |================================                                      |  50%
  |
  |==================================                                    |  53%
  |
  |====================================                                  |  57%
  |
  |======================================                                |  60%
  |
  |========================================                              |  63%
  |
  |==========================================                            |  67%
  |
  |============================================                          |  70%
  |
  |==============================================                        |  73%
  |
  |================================================                      |  77%
  |
  |==================================================                    |  80%
  |
  |====================================================                  |  83%
  |
  |======================================================                |  87%
  |
  |========================================================              |  90%
  |
  |==========================================================            |  93%
  |
  |============================================================          |  97%
  |
  |======================================================================| 100%
```

```
preds <- as.data.frame(h2o.predict(model, newdata=test.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(test$class, as.vector(preds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9396
```

```
#Tree 30, depth 16
model <- h2o.randomForest(y=T.dep, x=T.indep, training_frame = train.h2o,
                          ntrees = 30, max_depth = 16, seed = 1234)
```

```
## Warning in .h2o.startModelJob(algo, params, h2oRestApiVersion): Dropping bad and constant columns: [
```

```
##
  |
  |                                                                      |   0%
  |
  |==                                                                    |   3%
  |
  |====                                                                  |   7%
  |
  |======                                                                |  10%
  |
  |=========                                                             |  13%
  |
  |===========                                                           |  17%
  |
  |=============                                                         |  20%
  |
  |===============                                                       |  23%
  |
  |=================                                                     |  27%
  |
  |====================                                                  |  30%
  |
  |======================                                                |  33%
  |
  |========================                                              |  37%
  |
  |==========================                                            |  40%
  |
  |============================                                          |  43%
  |
  |==============================                                        |  47%
  |
  |================================                                      |  50%
  |
  |==================================                                    |  53%
  |
  |====================================                                  |  57%
  |
  |======================================                                |  60%
  |
  |========================================                              |  63%
  |
  |==========================================                            |  67%
  |
  |============================================                          |  70%
  |
  |==============================================                        |  73%
  |
```

```
  |=================================================                |  77%
  |
  |==================================================               |  80%
  |
  |===================================================              |  83%
  |
  |====================================================             |  87%
  |
  |=====================================================            |  90%
  |
  |======================================================           |  93%
  |
  |=======================================================          |  97%
  |
  |========================================================================| 100%
```

```
preds <- as.data.frame(h2o.predict(model, newdata=test.h2o))
```

```
##
  |
  |                                                                 |   0%
  |
  |========================================================================| 100%
```

```
confusionMatrix <- table(test$class, as.vector(preds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9659
```

##RandomForest with stretched bounding box

```
library(h2o)
newtrain = read.csv("mnist_train_stretch.csv", header = TRUE)
newtest = read.csv("mnist_test_stretch.csv", header = TRUE)
h2o.init(nthreads = -1, max_mem_size = '4g', ip = "127.0.0.1", port = 50001)
```

```
##  Connection successful!
##
## R is connected to the H2O cluster:
##     H2O cluster uptime:         5 minutes 41 seconds
##     H2O cluster timezone:       America/Los_Angeles
##     H2O data parsing timezone:  UTC
##     H2O cluster version:        3.26.0.2
##     H2O cluster version age:    2 months and 7 days
##     H2O cluster name:           H2O_started_from_R_amishukl_scx813
##     H2O cluster total nodes:    1
##     H2O cluster total memory:   3.56 GB
##     H2O cluster total cores:    8
##     H2O cluster allowed cores:  8
##     H2O cluster healthy:        TRUE
##     H2O Connection ip:          127.0.0.1
##     H2O Connection port:        50001
##     H2O Connection proxy:       NA
##     H2O Internal Security:      FALSE
##     H2O API Extensions:         Amazon S3, XGBoost, Algos, AutoML, Core V3, Core V4
##     R Version:                  R version 3.5.2 (2018-12-20)
```

```
newtrain$class = as.factor(newtrain$class)
newtest$class = as.factor(newtest$class)

newtrain.h2o <- as.h2o(newtrain)
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
newtest.h2o <- as.h2o(newtest)
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
nT.dep <- 401
nT.indep <- c(1:400)

#Tree 10, depth 4
newmodel <- h2o.randomForest(y=nT.dep, x=nT.indep, training_frame = newtrain.h2o, ntrees = 10, max_depth
```

```
##
  |
  |                                                                      |   0%
  |
  |================================                                      |  50%
  |
  |======================================================================| 100%
```

```
newpreds <- as.data.frame(h2o.predict(newmodel, newdata=newtest.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(newtest$class, as.vector(newpreds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.8442
```

```
#Tree 10, depth 8
newmodel <- h2o.randomForest(y=nT.dep, x=nT.indep, training_frame =
                             newtrain.h2o, ntrees = 10,
                             max_depth = 8, seed = 120)
```

```
##
  |
  |                                                                      |   0%
  |
  |=============                                                         |  20%
  |
```

```
  |=========================                                             |  40%
  |
  |=================================                                     |  60%
  |
  |=========================================                             |  80%
  |
  |=================================================                     |  90%
  |
  |======================================================================| 100%
```

```
newpreds <- as.data.frame(h2o.predict(newmodel, newdata=newtest.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(newtest$class, as.vector(newpreds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.94
```

```
#Tree 10, depth 16
newmodel <- h2o.randomForest(y=nT.dep, x=nT.indep, training_frame =
                             newtrain.h2o, ntrees = 10, max_depth = 16,
                             seed = 120)
```

```
##
  |
  |                                                                      |   0%
  |
  |======                                                                |  10%
  |
  |=============                                                         |  20%
  |
  |====================                                                  |  30%
  |
  |==========================                                            |  40%
  |
  |================================                                      |  50%
  |
  |=======================================                               |  60%
  |
  |===============================================                       |  70%
  |
  |==============================================                        |  80%
  |
  |=====================================================                 |  90%
  |
  |======================================================================| 100%
```

```
newpreds <- as.data.frame(h2o.predict(newmodel, newdata=newtest.h2o))
```

```
##
  |
```

```
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(newtest$class, as.vector(newpreds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9562
```

```
#Tree 20, depth 4
newmodel <- h2o.randomForest(y=nT.dep, x=nT.indep, training_frame =
                             newtrain.h2o, ntrees = 20,
                             max_depth = 4, seed = 120)
```

```
##
  |
  |                                                                      |   0%
  |
  |================                                                      |  25%
  |
  |================================                                      |  55%
  |
  |================================================                      |  75%
  |
  |======================================================================| 100%
```

```
newpreds <- as.data.frame(h2o.predict(newmodel, newdata=newtest.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(newtest$class, as.vector(newpreds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.8552
```

```
#Tree 20, depth 8
newmodel <- h2o.randomForest(y=nT.dep, x=nT.indep, training_frame =
                             newtrain.h2o, ntrees = 20,
                             max_depth = 8, seed = 120)
```

```
##
  |
  |                                                                      |   0%
  |
  |===                                                                   |   5%
  |
  |==========                                                            |  15%
  |
  |================                                                      |  25%
  |
  |=======================                                               |  35%
  |
```

```
  |=========================                                             |  40%
  |
  |===============================                                       |  50%
  |
  |====================================                                  |  60%
  |
  |==========================================                            |  65%
  |
  |================================================                      |  75%
  |
  |=====================================================                 |  80%
  |
  |============================================================          |  85%
  |
  |===================================================================   |  95%
  |
  |======================================================================| 100%
```

```
newpreds <- as.data.frame(h2o.predict(newmodel, newdata=newtest.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(newtest$class, as.vector(newpreds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9456
```

```
#Tree 20, depth 16
newmodel <- h2o.randomForest(y=nT.dep, x=nT.indep, training_frame =
                             newtrain.h2o, ntrees = 20,
                             max_depth = 16, seed = 120)
```

```
##
  |
  |                                                                      |   0%
  |
  |===                                                                   |   5%
  |
  |======                                                                |  10%
  |
  |==========                                                            |  15%
  |
  |=============                                                         |  20%
  |
  |================                                                      |  25%
  |
  |====================                                                  |  30%
  |
  |========================                                              |  35%
  |
  |=========================                                             |  40%
```

```
  |
  |============================                                          |  45%
  |
  |================================                                      |  50%
  |
  |====================================                                  |  55%
  |
  |========================================                              |  60%
  |
  |============================================                          |  65%
  |
  |================================================                      |  70%
  |
  |====================================================                  |  75%
  |
  |========================================================              |  80%
  |
  |============================================================          |  85%
  |
  |================================================================      |  90%
  |
  |====================================================================  |  95%
  |
  |======================================================================| 100%
```

```
newpreds <- as.data.frame(h2o.predict(newmodel, newdata=newtest.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(newtest$class, as.vector(newpreds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9652
```

```
#Tree 30, depth 4
newmodel <- h2o.randomForest(y=nT.dep, x=nT.indep, training_frame =
                             newtrain.h2o, ntrees = 30,
                             max_depth = 4, seed = 120)
```

```
##
  |
  |                                                                      |   0%
  |
  |===========                                                           |  17%
  |
  |====================                                                  |  30%
  |
  |==============================                                        |  47%
  |
  |========================================                              |  60%
  |
```

```
  |=================================================                 |  73%
  |
  |=============================================================     |  90%
  |
  |=================================================================| 100%
```

```
newpreds <- as.data.frame(h2o.predict(newmodel, newdata=newtest.h2o))
```

```
##
  |
  |                                                                 |   0%
  |
  |=================================================================| 100%
```

```
confusionMatrix <- table(newtest$class, as.vector(newpreds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.86
```

```
#Tree 30, depth 8
newmodel <- h2o.randomForest(y=nT.dep, x=nT.indep, training_frame =
                             newtrain.h2o, ntrees = 30,
                             max_depth = 8, seed = 120)
```

```
##
  |
  |                                                                 |   0%
  |
  |==                                                               |   3%
  |
  |======                                                           |  10%
  |
  |=========                                                        |  13%
  |
  |=============                                                    |  20%
  |
  |===============                                                  |  23%
  |
  |====================                                             |  30%
  |
  |======================                                           |  33%
  |
  |=========================                                        |  40%
  |
  |============================                                     |  43%
  |
  |==============================                                   |  47%
  |
  |==================================                               |  53%
  |
  |====================================                             |  57%
  |
  |========================================                         |  63%
  |
  |==========================================                       |  67%
```

```
  |
  |=============================================                    |  70%
  |
  |==============================================                   |  77%
  |
  |================================================                 |  80%
  |
  |=================================================                |  83%
  |
  |===================================================              |  90%
  |
  |====================================================             |  93%
  |
  |=====================================================================| 100%
```

```
newpreds <- as.data.frame(h2o.predict(newmodel, newdata=newtest.h2o))
```

```
##
  |
  |                                                                 |   0%
  |
  |=====================================================================| 100%
```

```
confusionMatrix <- table(newtest$class, as.vector(newpreds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9474
```

```
#Tree 30, depth 16
newmodel <- h2o.randomForest(y=nT.dep, x=nT.indep, training_frame =
                             newtrain.h2o, ntrees = 30,
                             max_depth = 16, seed = 120)
```

```
##
  |
  |                                                                 |   0%
  |
  |==                                                               |   3%
  |
  |====                                                             |   7%
  |
  |======                                                           |  10%
  |
  |=========                                                        |  13%
  |
  |===========                                                      |  17%
  |
  |=============                                                    |  20%
  |
  |===============                                                  |  23%
  |
  |=================                                                |  27%
  |
  |====================                                             |  30%
  |
```

```
  |======================                                                |  33%
  |
  |========================                                              |  37%
  |
  |==========================                                            |  40%
  |
  |============================                                          |  43%
  |
  |==============================                                        |  47%
  |
  |================================                                      |  50%
  |
  |==================================                                    |  53%
  |
  |====================================                                  |  57%
  |
  |======================================                                |  60%
  |
  |========================================                              |  63%
  |
  |==========================================                            |  67%
  |
  |============================================                          |  70%
  |
  |==============================================                        |  73%
  |
  |================================================                      |  77%
  |
  |==================================================                    |  80%
  |
  |====================================================                  |  83%
  |
  |======================================================                |  87%
  |
  |========================================================              |  90%
  |
  |==========================================================            |  93%
  |
  |============================================================          |  97%
  |
  |======================================================================| 100%
```

```
newpreds <- as.data.frame(h2o.predict(newmodel, newdata=newtest.h2o))
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

```
confusionMatrix <- table(newtest$class, as.vector(newpreds$predict))
accuracy <- sum(diag(confusionMatrix))/sum(confusionMatrix)
accuracy
```

```
## [1] 0.9678
```