

rsuite

Telco Customer Churn

Create project

```
msfz751@HPZB00K15XG4 R:\rsuite-projects\telco-customer-churn-tutorial  
$ rsuite proj start -n telco-customer-churn-tutorial
```

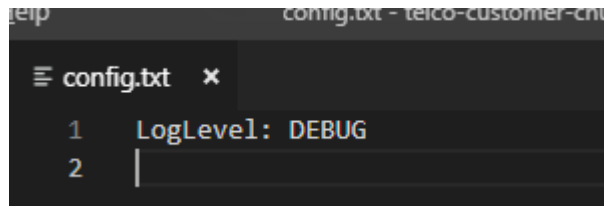
Add packages

- We add two packages

```
rsuite proj pkgadd -n externalpackages
```

```
rsuite proj pkgadd -n modelbuilder
```

config: Add debug level



The screenshot shows a code editor window with a dark theme. The title bar at the top reads "config.txt - telco-customer-ch". Below the title bar, the editor shows a file named "config.txt" with a close button. The content of the file is as follows:

```
1 LogLevel: DEBUG
2 |
```

externalpackages: Add dependencies

- Click to add Text

```
DESCRIPTION x
1 Package: externalpackages
2 Type: Package
3 Title: What the package does (short line)
4 Version: 0.1
5 Date: 2019-05-04
6 Author: msfz751
7 Maintainer: Who to complain to <yourfault@somewhere.net>
8 Description: More about what it does (maybe more than one line)
9 License: What license is it under?
10 Imports: logging,
11 | data.table
12 RoxygenNote: 6.1.1
13
```

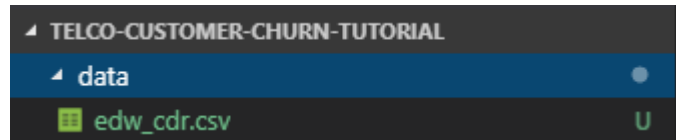
```
packages_import.R x
1 #-----
2 # modelbuilder
3 #
4 # Global package definitions and imports
5 #-----
6
7 #' @import logging
8 #' @import logging
9 NULL
10
```

modelbuilder: add dependencies

```
DESCRIPTION x
1 Package: modelbuilder
2 Type: Package
3 Title: What the package does (short line)
4 Version: 0.1
5 Date: 2019-05-04
6 Author: msfz751
7 Maintainer: Who to complain to <yourfault@somewhere.net>
8 Description: More about what it does (maybe more than one line)
9 License: What license is it under?
10 Imports: logging,
11 |       h2o
12 RoxygenNote: 6.1.1
13 |
```

```
packages_import.R x
1 #-----
2 # modelbuilder
3 #
4 # Global package definitions and imports
5 #-----
6
7 #' @import logging
8 #' @import h2o
9 NULL
10
```

Add data files



modelbuilder: Add script to find best model

- We add only one script to the package

```
find_best_model.R x
1  #'@export
2  find_best_classifier_model <- function(model_quality_measure = "AUC", ...) {
3    h2o_grid <- h2o.grid(...)
4
5    best_model <- list(
6      model = h2o.getModel(h2o.getGrid(h2o_grid@grid_id,
7                                     sort_by = model_quality_measure,
8                                     decreasing = TRUE)@model_ids[[1]]))
9
10   best_model$threshold <- h2o.find_threshold_by_max_metric(h2o.performance(best_model,
11                                                                           xval = TRUE,
12                                                                           "min_per_class_accuracy"))
13   return(best_model)
14 }
15
```


project: add 1st script to build model

- The script will run from the project.
- It will run the model calling a function from modelbuilder.

```
build_telco_churn_model.R
12 config <- load_config()
13 args <- args_parser()
14
15 #####
16
17 data_path <- file.path(script_path, "../data")
18 export_path <- file.path(script_path, "../export")
19
20 library(externalpackages)
21 library(modelbuilder)
22
23 loginfo("--> Loading data...")
24 all_data <- fread(file.path(data_path, "edw_cdr.csv"))
25 all_data <- all_data[, !c("month", "year"), with = FALSE]
26 all_data <- all_data[complete.cases(all_data)]
27 all_data <- all_data[!duplicated(all_data)]
28
29 loginfo("--> loaded %s rows", nrow(all_data))
30
31 set.seed(1234)
32
33 loginfo("--> Cleaning&transforming data...")
34 all_data[, ind := factor(sample(0:1, size = .N, replace = TRUE, prob = c(0.3, 0.7)),
35                           levels = 0:1,
36                           labels = c("Test", "Train"))]
37 all_data[, churn := factor(ifelse(churn == 1, "churn", "nochurn"))]
38
39 loginfo("--> done")
40
41 nthreads <- as.integer(args$get("nthreads", required = FALSE, default = -1))
42 max_mem_size <- args$get("max_mem", required = FALSE, default = "4g")
```

terminal: Install dependencies and build project

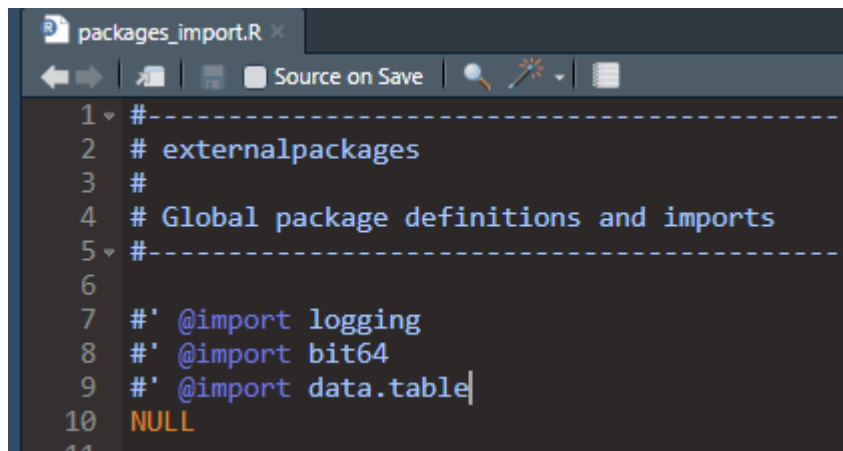
```
$ rsuite proj depsinst
2019-05-04 18:17:47 INFO:rsuite:Detecting repositories (for R 3.5)...
2019-05-04 18:17:48 INFO:rsuite:Will look for dependencies in ...
2019-05-04 18:17:48 INFO:rsuite:.          MRAN#1 = https://mran.microsoft.com/snapshot/2019-05-04
2019-05-04 18:17:48 INFO:rsuite:Collecting project dependencies (for R 3.5)...
2019-05-04 18:17:48 INFO:rsuite:Resolving dependencies (for R 3.5)...
2019-05-04 18:17:49 INFO:rsuite:Following installed packages will be updated: externalpackag
2019-05-04 18:17:49 INFO:rsuite:Detected 3 dependencies to install. Installing...
2019-05-04 18:18:00 INFO:rsuite:All dependencies successfully installed.
```

```
$ rsuite proj build
2019-05-04 18:18:27 INFO:rsuite:Installing externalpackages (for R 3.5) ...
2019-05-04 18:18:27 INFO:rsuite:... not changed; keeping previously installed package.
2019-05-04 18:18:27 INFO:rsuite:Installing modelbuilder (for R 3.5) ...
2019-05-04 18:18:27 INFO:rsuite:... not changed; keeping previously installed package.
2019-05-04 18:18:27 INFO:rsuite:Successfully installed 2 packages
```

- The execution stops with a warning about data.table requiring bit64.

externalpackages: add package bit64

- the package **bit64** is needed by **data.table**.



A screenshot of an R script editor window titled 'packages_import.R'. The editor shows a script with the following content:

```
1 #-----  
2 # externalpackages  
3 #  
4 # Global package definitions and imports  
5 #-----  
6  
7 #' @import logging  
8 #' @import bit64  
9 #' @import data.table  
10 NULL  
11
```

terminal: Install dependencies and build project

- Click to add Text

```
msfz751@HPZB00K15XG4 R:\rsuite-projects\telco-customer-churn-tutorial
$ rsuite proj depsinst
2019-05-04 18:25:44 INFO:rsuite:Detecting repositories (for R 3.5)...
2019-05-04 18:25:44 INFO:rsuite:Will look for dependencies in ...
2019-05-04 18:25:44 INFO:rsuite:..          MRAN#1 = https://mran.microsoft.com/snapshot/2019-05-04
2019-05-04 18:25:44 INFO:rsuite:Collecting project dependencies (for R 3.5)...
2019-05-04 18:25:44 INFO:rsuite:Resolving dependencies (for R 3.5)...
2019-05-04 18:25:46 INFO:rsuite:Following installed packages will be updated: externalpac
Rcpp
2019-05-04 18:25:46 INFO:rsuite:Detected 2 dependencies to install. Installing...
2019-05-04 18:25:50 INFO:rsuite:All dependencies successfully installed.
```

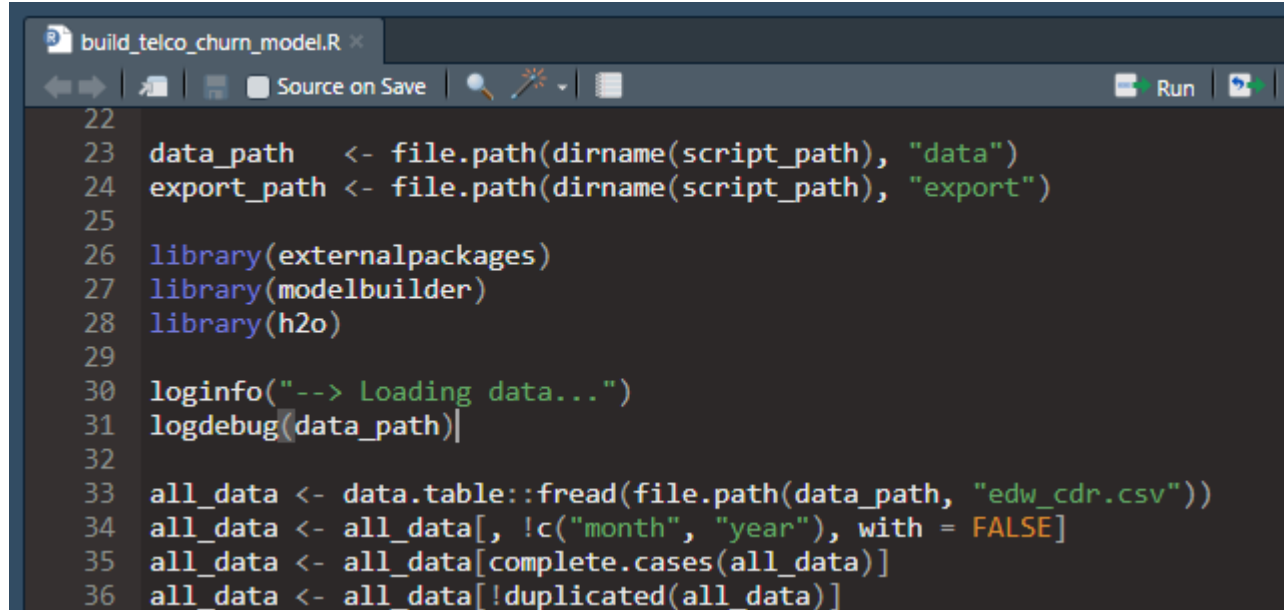
terminal: Run model

```
$ Rscript R/build_telco_churn_model.R
2019-05-04 18:27:03 INFO:--> Loading data...
2019-05-04 18:27:03 DEBUG::R:/rsuite-projects/telco-customer-churn-tutorial/data
2019-05-04 18:27:03 INFO:--> loaded 20467 rows
```

- During the execution the script stops because h2o is not loaded in the global environment. We will add `library(h2o)`
- The alternative is calling `h2o.init` like this: `h2o::h2o.init()`

project: Add library to load package h2o

- Since we are calling `h2o.init` from the project body, we should load h2o also from the project.



```
build_telco_churn_model.R x
Source on Save
Run

22
23 data_path <- file.path(dirname(script_path), "data")
24 export_path <- file.path(dirname(script_path), "export")
25
26 library(externalpackages)
27 library(modelbuilder)
28 library(h2o)
29
30 loginfo("--> Loading data...")
31 logdebug(data_path)
32
33 all_data <- data.table::fread(file.path(data_path, "edw_cdr.csv"))
34 all_data <- all_data[, !c("month", "year"), with = FALSE]
35 all_data <- all_data[complete.cases(all_data)]
36 all_data <- all_data[!duplicated(all_data)]
```

terminal: Run model

- Running the model takes about 7 minutes.

```
$ Rscript R/build_telco_churn_model.R

-----

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'
```

terminal: Results

- We get accuracy, recall and AUC.

```
[1] 0
|=====| 100%
|=====| 100%
2019-05-04 18:42:04 INFO:--> Datasets imported into H2O cluster
|=====| 100%
[1] "R:\\rsuite-projects\\telco-customer-churn-tutorial\\export\\gbm_grid_model_7"
2019-05-04 18:51:33 INFO:--> Best model exported into export folder
2019-05-04 18:51:33 INFO:--> Best model with AUC=0.919339425749523
2019-05-04 18:51:33 INFO:--> Threshold for min per class accuracy metric = 0.999754395090411
|=====| 100%
2019-05-04 18:51:34 INFO:--> Scoring test datasets done
2019-05-04 18:51:34 INFO:Test accuracy: 0.835
2019-05-04 18:51:34 INFO:Test precision: 0.344
2019-05-04 18:51:34 INFO:Test recall: 0.887
2019-05-04 18:51:34 INFO:Test AUC: 0.952
msf-751QNP7BQK15XG4 R:\rsuite-projects\telco-customer-churn-tutorial
```