R version 3.3.3 (2017-03-06) -- "Another Canoe"

Copyright (C) 2017 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> #TRUE: server settings / FALSE: local settings

> isServerRun = FALSE

> model.name <- "LDA - Feature Selection"

> model.formula <-as.formula("factor(upselling)~.")

>

> print(sprintf("model: %s formula: %s", model.name, deparse(model.formula)))

[1] "model: LDA - Feature Selection formula: factor(upselling) ~ ."

>

> #########################################

> ############## Drive Config #############

> #########################################

> if(isServerRun){

+ setwd('/host/dsm1/fmare001/stats/svm/deliverables')

+ }else{

+ #setwd('C:/Users/audrey.ekuban/dev/goldsmiths/mlsdm/assignment3')

+ #setwd('C:/Users/john/dev/goldsmiths/mlsdm/assignment3')

+ setwd('C:/Users/Fred/Desktop/Studies/MSc-DataScience/Statistical Learning/Assignments/Assignment3/deliverables')

+ }

>

> #########################################

> ########### Load Dependencies ###########

> #########################################

> source("init\_data.r")

Loading required package: caret

Loading required package: lattice

Loading required package: ggplot2

Loading required package: DMwR

Loading required package: grid

> source("exploratory\_functions.r")

Loading required package: dplyr

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

filter, lag

The following objects are masked from ‘package:base’:

intersect, setdiff, setequal, union

Loading required package: reshape2

> source("pre\_processing\_functions.r")

> source("feat\_selection.r")

randomForest 4.6-12

Type rfNews() to see new features/changes/bug fixes.

Attaching package: ‘randomForest’

The following object is masked from ‘package:dplyr’:

combine

The following object is masked from ‘package:ggplot2’:

margin

> require(randomForest)

>

> forceReloadPreCanned1 = TRUE

> if (forceReloadPreCanned1) {

+ #List the pre-processing functions

+ model.preProcessingFunctions <- c(

+ convert\_to\_factors,

+ drop\_na\_cols,

+ remove\_correlated\_predictors,

+ convert\_NAs\_to\_level,

+ remove\_linear\_dependencies,

+ bin\_negative\_levels\_churn,

+ keep\_top\_10\_levels,

+ impute\_data

+ )

+ # Need to do these up-front, otherwise we might end up with mis-matched

+ # levels between training and test folds in the cross-validation loop.

+ # Shouldn't introduce any bias since nothing is being imputed.

+ model.data <- apply\_pre\_processing(train, model.preProcessingFunctions)

+ write.csv(model.data, file = "train.bin.neg.top10.upselling.csv", row.names = FALSE)

+ } else {

+ #reload from file

+ model.data <- read.csv("train.bin.neg.top10.upselling.csv", stringsAsFactors = FALSE)

+ model.data <- convert\_to\_factors(model.data)

+ }

[1] "Number of removed linearly dependent col(s): 0"

>

> random\_forest\_feature\_selection(model.formula, model.data,500)

-1 1 MeanDecreaseAccuracy MeanDecreaseGini

V84 80.0639947 139.785083728 117.9745790 1124.896829

V101 37.7243434 45.622246309 49.0131965 182.456195

V14 23.6578558 7.724051038 25.3310446 107.640339

V155 17.1019219 0.006022873 16.7462901 106.939125

V111 0.4104939 2.775427262 1.3634383 101.957236

V70 11.5612909 -0.250573796 11.1159292 100.196848

V222 25.2643725 -0.948913210 25.7863515 99.717698

V192 26.3733059 0.263676948 26.3101572 99.114015

V141 24.7402084 -3.850432215 24.9884348 87.379857

V161 27.0046285 -3.958914806 27.1140696 86.860121

V191 25.7682405 -3.877923315 25.3151729 82.655551

V55 30.1835721 -8.565758819 30.2574518 80.080482

V156 6.3187242 -1.131445699 5.6256756 78.966269

V91 24.0193883 -3.941621729 24.4687729 77.755641

V37 16.5847582 -4.045787105 16.8302406 76.375828

V116 21.2214619 1.870144263 21.8060023 76.019916

V42 19.5053311 -7.400616286 19.5329504 75.572494

V134 22.6094769 -2.121794792 22.9542789 75.518938

V189 21.6711340 -6.696429596 21.1069025 71.115267

V96 20.4802239 -4.178329974 20.0212207 68.870745

V1 25.1674676 -3.388000124 25.2777919 68.474461

V121 18.2213199 -1.460217384 18.1387476 67.038085

V166 28.7652543 -7.854682032 28.7631352 65.890675

V124 28.3439232 -9.315845576 27.7788925 65.569100

V81 16.5176831 -5.788985641 16.3620312 65.489848

V151 27.6267117 -7.300175158 27.4788771 64.770253

V75 19.0557333 -1.638078268 19.1470506 63.809633

V206 24.3361120 -2.506900020 24.3115058 58.378400

V197 27.1924353 -4.849814558 26.6481601 55.745869

V165 -0.3819260 -0.537881721 -0.5485696 54.940563

V202 21.8774573 -5.629937105 21.8841582 53.436084

V150 13.1986503 -3.190311213 13.1975373 50.861781

V15 18.8536142 -0.932041808 18.9688273 49.891702

V119 22.9682520 -4.594644799 22.3292309 44.332867

V21 18.4885213 -1.465306149 18.3229075 43.023572

V71 22.6162494 -3.600415079 22.7762552 41.796519

V67 15.3733820 -0.287594610 15.2016059 41.418291

V29 22.4594709 -7.730953694 21.9494637 40.076310

V190 20.7580855 -7.691624598 20.2042501 39.373488

V138 13.4406663 1.039273572 13.2643447 36.974562

V58 9.8893270 -3.636561116 9.1787014 36.321383

V168 5.7660109 34.174246735 23.7159878 35.717504

V154 12.5857810 -6.534588479 12.2102747 32.507588

V107 11.4037430 -1.418065852 11.2940777 29.488693

V115 10.5858919 -2.102584065 10.6383229 29.104532

appetency 26.5671767 21.432616404 27.7350218 27.991419

V114 6.2886842 -2.116859680 6.1248306 27.253069

V118 17.9668532 5.099668655 18.8566243 27.032800

V143 4.1503567 0.860110444 4.2407361 23.977202

V90 17.9659580 -7.689856470 17.2032629 23.882431

V187 17.8940201 -8.662944304 16.7793296 23.457143

V93 15.3826144 -7.141779419 14.8060468 23.218768

V95 12.4033681 2.255270579 12.8758911 22.985942

V204 15.7589001 -6.499085894 15.1979102 20.954009

V149 12.7202296 7.145903859 13.6298577 19.852805

V212 9.1038167 -3.056338394 8.7780245 17.928533

V112 9.2480769 -3.708705899 9.0092676 17.243653

V79 8.3224159 0.931353031 8.3476176 16.875518

V224 15.1740214 -2.071399878 14.8323462 15.583301

V171 6.9068128 1.527694406 6.8289636 14.893370

V30 9.6677182 -2.703900826 8.0655323 13.594566

churn 9.4659200 20.956291115 24.4322721 13.130734

V230 11.4411118 -0.636887951 11.5346852 12.984308

V11 4.4324388 -4.197403672 3.2996984 11.864288

V196 3.5239318 -0.848194003 2.9708824 8.902878

V94 11.5817796 -1.921091380 11.6388474 6.401528

V113 10.8702635 -1.227701429 11.0205575 6.052292

V128 4.9704251 10.151999703 10.0413764 4.741697

V125 4.1275338 0.073777137 3.8749786 4.425483

V219 5.4936089 0.397298188 5.6490194 2.387522

V73 0.9398035 -1.620561400 0.4854529 1.121520

