Maching Learning in R using mIR

Surag Gupta

August 20, 2017

About me

- Senior Analyst, Assortment Optimization
- ▶ MS Business Analytics, University of Minnesota (Class of 2017)
- which means I'm a relatively new team member
- Worked as a Decision Scientist for 3 years prior to my graduate program
- R Enthusiast (This would be a good time for you to notice my t-shirt, if you haven't already)

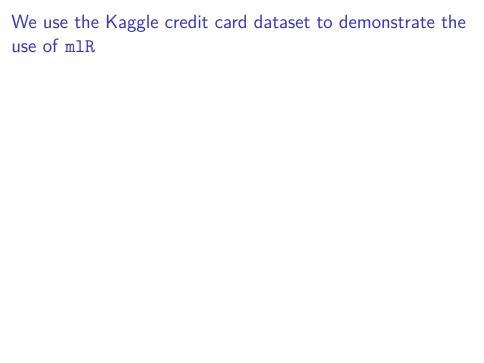


Those who use Python for data science tasks shall not throw stuff at me during the session

What is mIR?

mIR is an R package developed to simplify basic and advanced data science tasks, such as:

- Data preparation
- Feature Engineering
- ▶ Training Models
- Validating models
- Predictions



Load packages

1	gbm	kknn	caret	Hmisc	##
-	TRUE	TRUE	TRUE	TRUE	##
parallel	parallel	mlr	lightgbm	xgboost	##
-	TRUE	TRUE	TRUE	TRUE	##
			tidyverse	data.table	##
			TRUE	TRUE	##

Import the dataset

```
#Use fread() from the data.table package to read
df <- fread("creditcard.csv")</pre>
df <- df %>%
  data.frame() %>%
  mutate(Class = as.factor(Class)) #Convert target variable
head(df)
##
     Time
                  V1
                               V2
                                          V3
                                                      V4
```

0 -1.3598071 -0.07278117 2.5363467 1.3781552 -0.33 ## 1 ## 2

0 1.1918571 0.26615071 0.1664801 0.4481541 0.00 ## 3 1 -1.3583541 -1.34016307 1.7732093 0.3797796 -0.50 ## 4 1 -0.9662717 -0.18522601 1.7929933 -0.8632913 -0.0 ## 5 2 -1.1582331 0.87773675 1.5487178 0.4030339 -0.40

6 2 -0.4259659 0.96052304 1.1411093 -0.1682521 0.49

٧7 V8 V10 ۷9 ## 1 0.23959855 0.09869790 0.3637870 0.09079417 -0.551