Projektbericht2

Ferdinand Bubeck und Johannes Bubeck

25 3 2021

#import libraries  
library(readr)  
library(dplyr)  
library(ggplot2)  
library(tidyverse)  
  
datTest <- read\_csv("aug\_test.csv")  
  
datTrain <- read\_csv("aug\_train.csv")  
  
#Data Preparation  
datTrain\_casted <- datTrain %>%   
 select(-enrollee\_id, -city) %>%  
 mutate(city\_development\_index = as.numeric(city\_development\_index),  
 target = as.factor(target)) %>%   
 as.data.frame() %>%  
 mutate\_if(is.character, as.factor) #converting all variables with type=char to factor  
  
  
datTrain\_notImputed <- datTrain\_casted  
  
datTrain\_NaReduced <- na.omit(datTrain\_notImputed)  
  
#NAs  
detectNA <- function(x){sum(is.na(x))/length(x)\*100}  
apply(datTrain\_NaReduced, 2, detectNA)

## city\_development\_index gender relevent\_experience   
## 0 0 0   
## enrolled\_university education\_level major\_discipline   
## 0 0 0   
## experience company\_size company\_type   
## 0 0 0   
## last\_new\_job training\_hours target   
## 0 0 0

#Test / Training Split  
set.seed(304)  
trainingRows2 <- sort(sample(nrow(datTrain\_NaReduced), nrow(datTrain\_NaReduced)\*.7))  
  
Train\_NaReduced <- datTrain\_NaReduced[trainingRows2,]  
Test\_NaReduced <- datTrain\_NaReduced[-trainingRows2,]  
  
set.seed(7)  
glm.fit\_NotImputed <- glm(target ~ ., family = binomial, Train\_NaReduced)  
summary(glm.fit\_NotImputed)

##   
## Call:  
## glm(formula = target ~ ., family = binomial, data = Train\_NaReduced)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.9987 -0.4723 -0.3809 -0.3214 2.6383   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 5.324368 0.621200 8.571 < 2e-16  
## city\_development\_index -8.083843 0.315892 -25.590 < 2e-16  
## genderMale -0.044523 0.137908 -0.323 0.746814  
## genderOther 0.675955 0.398853 1.695 0.090124  
## relevent\_experienceNo relevent experience 0.181078 0.117742 1.538 0.124068  
## enrolled\_universityno\_enrollment -0.045326 0.121305 -0.374 0.708663  
## enrolled\_universityPart time course -0.184680 0.187970 -0.983 0.325853  
## education\_levelMasters -0.043099 0.089034 -0.484 0.628334  
## education\_levelPhd 0.196926 0.259206 0.760 0.447418  
## major\_disciplineBusiness Degree 0.093910 0.560552 0.168 0.866953  
## major\_disciplineHumanities 0.339558 0.497487 0.683 0.494894  
## major\_disciplineNo Major 0.411614 0.575675 0.715 0.474602  
## major\_disciplineOther -0.152067 0.553058 -0.275 0.783350  
## major\_disciplineSTEM 0.394302 0.452689 0.871 0.383742  
## experience>20 -1.086418 0.316136 -3.437 0.000589  
## experience1 0.250043 0.375504 0.666 0.505483  
## experience10 -0.936664 0.326816 -2.866 0.004157  
## experience11 -1.005113 0.350262 -2.870 0.004110  
## experience12 -1.245351 0.371686 -3.351 0.000807  
## experience13 -0.775835 0.374835 -2.070 0.038471  
## experience14 -0.749282 0.360251 -2.080 0.037536  
## experience15 -0.863632 0.351664 -2.456 0.014055  
## experience16 -1.118979 0.373322 -2.997 0.002723  
## experience17 -1.016150 0.399502 -2.544 0.010973  
## experience18 -1.242205 0.454256 -2.735 0.006246  
## experience19 -1.316183 0.441330 -2.982 0.002861  
## experience2 -0.867174 0.340529 -2.547 0.010879  
## experience20 -1.314140 0.612151 -2.147 0.031813  
## experience3 -0.646738 0.320822 -2.016 0.043812  
## experience4 -0.727593 0.320538 -2.270 0.023213  
## experience5 -0.704407 0.314003 -2.243 0.024877  
## experience6 -0.804031 0.319316 -2.518 0.011803  
## experience7 -0.780953 0.323626 -2.413 0.015816  
## experience8 -0.792071 0.332284 -2.384 0.017139  
## experience9 -0.961829 0.331265 -2.904 0.003690  
## company\_size10/49 0.355596 0.173953 2.044 0.040932  
## company\_size100-500 0.130646 0.166150 0.786 0.431683  
## company\_size1000-4999 0.357188 0.188932 1.891 0.058682  
## company\_size10000+ 0.366862 0.170657 2.150 0.031579  
## company\_size50-99 0.162006 0.160005 1.013 0.311295  
## company\_size500-999 0.269732 0.203739 1.324 0.185533  
## company\_size5000-9999 0.301206 0.233584 1.289 0.197225  
## company\_typeFunded Startup -0.074665 0.224272 -0.333 0.739195  
## company\_typeNGO -0.258444 0.278912 -0.927 0.354127  
## company\_typeOther 0.559782 0.424779 1.318 0.187564  
## company\_typePublic Sector 0.243295 0.245440 0.991 0.321559  
## company\_typePvt Ltd -0.050304 0.189342 -0.266 0.790486  
## last\_new\_job1 -0.107622 0.121770 -0.884 0.376796  
## last\_new\_job2 0.091014 0.135640 0.671 0.502224  
## last\_new\_job3 0.183896 0.170551 1.078 0.280925  
## last\_new\_job4 0.214406 0.176159 1.217 0.223558  
## last\_new\_jobnever 0.087343 0.193051 0.452 0.650957  
## training\_hours -0.001112 0.000648 -1.716 0.086185  
##   
## (Intercept) \*\*\*  
## city\_development\_index \*\*\*  
## genderMale   
## genderOther .   
## relevent\_experienceNo relevent experience   
## enrolled\_universityno\_enrollment   
## enrolled\_universityPart time course   
## education\_levelMasters   
## education\_levelPhd   
## major\_disciplineBusiness Degree   
## major\_disciplineHumanities   
## major\_disciplineNo Major   
## major\_disciplineOther   
## major\_disciplineSTEM   
## experience>20 \*\*\*  
## experience1   
## experience10 \*\*   
## experience11 \*\*   
## experience12 \*\*\*  
## experience13 \*   
## experience14 \*   
## experience15 \*   
## experience16 \*\*   
## experience17 \*   
## experience18 \*\*   
## experience19 \*\*   
## experience2 \*   
## experience20 \*   
## experience3 \*   
## experience4 \*   
## experience5 \*   
## experience6 \*   
## experience7 \*   
## experience8 \*   
## experience9 \*\*   
## company\_size10/49 \*   
## company\_size100-500   
## company\_size1000-4999 .   
## company\_size10000+ \*   
## company\_size50-99   
## company\_size500-999   
## company\_size5000-9999   
## company\_typeFunded Startup   
## company\_typeNGO   
## company\_typeOther   
## company\_typePublic Sector   
## company\_typePvt Ltd   
## last\_new\_job1   
## last\_new\_job2   
## last\_new\_job3   
## last\_new\_job4   
## last\_new\_jobnever   
## training\_hours .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 5591 on 6267 degrees of freedom  
## Residual deviance: 4528 on 6215 degrees of freedom  
## AIC: 4634  
##   
## Number of Fisher Scoring iterations: 5

#Variable Importance  
library(caret)  
varImp(glm.fit\_NotImputed)

## Overall  
## city\_development\_index 25.5904920  
## genderMale 0.3228426  
## genderOther 1.6947465  
## relevent\_experienceNo relevent experience 1.5379202  
## enrolled\_universityno\_enrollment 0.3736527  
## enrolled\_universityPart time course 0.9825005  
## education\_levelMasters 0.4840737  
## education\_levelPhd 0.7597269  
## major\_disciplineBusiness Degree 0.1675303  
## major\_disciplineHumanities 0.6825461  
## major\_disciplineNo Major 0.7150113  
## major\_disciplineOther 0.2749563  
## major\_disciplineSTEM 0.8710221  
## experience>20 3.4365486  
## experience1 0.6658871  
## experience10 2.8660302  
## experience11 2.8696027  
## experience12 3.3505449  
## experience13 2.0698000  
## experience14 2.0798909  
## experience15 2.4558421  
## experience16 2.9973539  
## experience17 2.5435425  
## experience18 2.7345941  
## experience19 2.9823111  
## experience2 2.5465494  
## experience20 2.1467581  
## experience3 2.0158814  
## experience4 2.2699151  
## experience5 2.2433137  
## experience6 2.5179805  
## experience7 2.4131313  
## experience8 2.3837170  
## experience9 2.9035007  
## company\_size10/49 2.0442135  
## company\_size100-500 0.7863156  
## company\_size1000-4999 1.8905680  
## company\_size10000+ 2.1497053  
## company\_size50-99 1.0125076  
## company\_size500-999 1.3239099  
## company\_size5000-9999 1.2894976  
## company\_typeFunded Startup 0.3329191  
## company\_typeNGO 0.9266146  
## company\_typeOther 1.3178204  
## company\_typePublic Sector 0.9912600  
## company\_typePvt Ltd 0.2656791  
## last\_new\_job1 0.8838155  
## last\_new\_job2 0.6709950  
## last\_new\_job3 1.0782436  
## last\_new\_job4 1.2171207  
## last\_new\_jobnever 0.4524336  
## training\_hours 1.7158726