Auto Isurance Assignment

Prasanna_Rao_Sec 58

Monday, February 16, 2015

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(rattle)
```

```
## Rattle: A free graphical interface for data mining with R.
## Version 3.4.1 Copyright (c) 2006-2014 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```
library(ipred)
  library(lava)
```

```
## lava version 1.3
```

```
library(rpart)
library(tree)
library(caret)
```

```
## Loading required package: lattice
## Loading required package: ggplot2
##
## Attaching package: 'ggplot2'
##
## The following object is masked from 'package:lava':
##
##     %+%
```

```
library(randomForest)
```

```
## randomForest 4.6-10
## Type rfNews() to see new features/changes/bug fixes.
```

```
infile1<-read.csv('C:/Prasanna Krishna/Analytics/MS/411/Unit2/cars.csv',stringsAsFactors=FALSE,na.stri
infile1<-data.frame(infile1)
infile2<-infile1



infile2$INCOME <- as.numeric(gsub('\\$|,',"",infile2$INCOME))
infile2$BLUEBOOK <- as.numeric(gsub('\\$|,',"",infile2$BLUEBOOK))
infile2$HOME_VAL <- as.numeric(gsub('\\$|,',"",infile2$HOME_VAL))
```

```
infile2$OLDCLAIM <- as.numeric(gsub('\\$|,',"",infile2$OLDCLAIM))
```

```
dummies_SEX<-data.frame(dummies_SEX<-predict(dummyVars(~ SEX, data=infile2),newdata=infile2))
dummies_PARENT1<-data.frame(dummies_PARENT1<-predict(dummyVars(~ PARENT1, data=infile2),newdata=infil
dummies_MSTATUS<-data.frame(predict(dummyVars(~ MSTATUS, data=infile2),newdata=infile2))
dummies_EDUCATION<-data.frame(predict(dummyVars(~ EDUCATION, data=infile2),newdata=infile2))
dummies_JOB<-data.frame(predict(dummyVars(~ JOB, data=infile2),newdata=infile2))
dummies_CAR_USE<-data.frame(predict(dummyVars(~ CAR_USE, data=infile2),newdata=infile2))
dummies_CAR_TYPE<-data.frame(predict(dummyVars(~ CAR_TYPE, data=infile2),newdata=infile2))
dummies_RED_CAR<-data.frame(predict(dummyVars(~ RED_CAR, data=infile2),newdata=infile2))
dummies_REVOKED<-data.frame(predict(dummyVars(~ MSTATUS, data=infile2),newdata=infile2))
dummies_URBANICITY<-data.frame(predict(dummyVars(~ URBANICITY, data=infile2),newdata=infile2))
```

```
 infile3<-cbind(infile2,dummies_SEX,dummies_PARENT1,dummies_MSTATUS,dummies_EDUCATION,dummies_JOB,du
              dummies_RED_CAR,dummies_REVOKED,dummies_URBANICITY)
```

```
infile4<-infile3[ ,c(2,4,5,7,8,10,17,25,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,
```

```
preProc <- preProcess(method="bagImpute",infile4)

infile6 <- predict(preProc, infile4)

 infile2$YOJ<-infile6$YOJ
 infile2$INCOME<-infile6$INCOME
 infile2$HOME_VAL<-infile6$HOME_VAL
 infile2$CAR_AGE<-infile6$CAR_AGE
 infile2$AGE<-infile6$AGE

infile5<-infile2

infile5$PARENT1<-as.factor(infile5$PARENT1)
infile5$MSTATUS<-as.factor(infile5$MSTATUS)
infile5$SEX<-as.factor(infile5$SEX)
infile5$JOB<-as.factor(infile5$JOB)
infile5$CAR_USE<-as.factor(infile5$CAR_USE)
infile5$CAR_TYPE<-as.factor(infile5$CAR_TYPE)
infile5$EDUCATION<-as.factor(infile5$EDUCATION)
infile5$RED_CAR<-as.factor(infile5$RED_CAR)

infile5$REVOKED<-as.factor(infile5$REVOKED)
infile5$URBANICITY<-as.factor(infile5$URBANICITY)
 infile5<-cbind(infile5,dummies_JOB )
  infile61<-infile5[which(infile2$TARGET_FLAG ==1),]
   infile61<-infile61[,c(-1,-2,-14)]
  str(infile5)
```

```
## 'data.frame':    8161 obs. of  34 variables:
##  $ INDEX           : int  1 2 4 5 6 7 8 11 12 13 ...
##  $ TARGET_FLAG     : int  0 0 0 0 0 1 0 1 1 0 ...
##  $ TARGET_AMT      : num  0 0 0 0 0 ...
##  $ KIDSDRIV        : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ AGE             : num  60 43 35 51 50 34 54 37 34 50 ...
##  $ HOMEKIDS        : int  0 0 1 0 0 1 0 2 0 0 ...
##  $ YOJ             : num  11 11 10 14 12 ...
##  $ INCOME          : num  67349 91449 16039 108586 114986 ...
##  $ PARENT1         : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 2 1 1 1 1 ...
##  $ HOME_VAL        : num  0 257252 124191 306251 243925 ...
##  $ MSTATUS         : Factor w/ 2 levels "Yes","z_No": 2 2 1 1 1 2 1 1 2 2 ...
##  $ SEX             : Factor w/ 2 levels "M","z_F": 1 1 2 1 2 2 2 1 2 1 ...
##  $ EDUCATION       : Factor w/ 5 levels "<High School",..: 4 5 5 1 4 2 1 2 2 2 ...
##  $ JOB             : Factor w/ 8 levels "Clerical","Doctor",..: 6 8 1 8 2 8 8 8 1 6 ...
##  $ TRAVTIME        : int  14 22 5 32 36 46 33 44 34 48 ...
##  $ CAR_USE         : Factor w/ 2 levels "Commercial","Private": 2 1 2 2 2 1 2 1 2 1 ...
##  $ BLUEBOOK        : num  14230 14940 4010 15440 18000 ...
##  $ TIF             : int  11 1 4 7 1 1 1 1 1 7 ...
##  $ CAR_TYPE        : Factor w/ 6 levels "Minivan","Panel Truck",..: 1 1 6 1 6 4 6 5 6 5 ...
##  $ RED_CAR         : Factor w/ 2 levels "no","yes": 2 2 1 2 1 1 1 2 1 1 ...
##  $ OLDCLAIM        : num  4461 0 38690 0 19217 ...
##  $ CLM_FREQ        : int  2 0 2 0 2 0 0 1 0 0 ...
##  $ REVOKED         : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 1 2 1 1 ...
##  $ MVR_PTS         : int  3 0 3 0 3 0 0 10 0 1 ...
##  $ CAR_AGE         : num  18 1 10 6 17 7 1 7 1 17 ...
##  $ URBANICITY      : Factor w/ 2 levels "Highly Urban/ Urban",..: 1 1 1 1 1 1 1 1 1 2 ...
##  $ JOBClerical     : num  0 0 1 0 0 0 0 0 1 0 ...
##  $ JOBDoctor       : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ JOBHome.Maker   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ JOBLawyer       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ JOBManager      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ JOBProfessional : num  1 0 0 0 0 0 0 0 0 1 ...
##  $ JOBStudent      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ JOBz_Blue.Collar: num  0 1 0 1 0 1 1 1 0 0 ...
```

```
   infile5<-infile5[,c(-1,-3,-14)]
```

```
 fit <- glm(TARGET_FLAG ~ . ,data=infile5,family=binomial())
 fit1<-lm(TARGET_AMT ~.,data=infile61)

infile11<-read.csv('C:/Prasanna Krishna/Analytics/MS/411/Unit2/logit_insurance_test.csv',stringsAsFacto
infile11<-data.frame(infile11)
infile21<-infile11


 infile21$INCOME  <- as.numeric(gsub('\\$|,',"",infile21$INCOME))
 infile21$BLUEBOOK <- as.numeric(gsub('\\$|,',"",infile21$BLUEBOOK))
 infile21$HOME_VAL <- as.numeric(gsub('\\$|,',"",infile21$HOME_VAL))
 infile21$OLDCLAIM <- as.numeric(gsub('\\$|,',"",infile21$OLDCLAIM))
```

```
dummies1_SEX<-data.frame(dummies1_SEX<-predict(dummyVars(~ SEX, data=infile21),newdata=infile21))
  dummies1_PARENT1<-data.frame(dummies1_PARENT1<-predict(dummyVars(~ PARENT1, data=infile21),newdata=in
  dummies1_MSTATUS<-data.frame(predict(dummyVars(~ MSTATUS, data=infile21),newdata=infile21))
  dummies1_EDUCATION<-data.frame(predict(dummyVars(~ EDUCATION, data=infile21),newdata=infile21))
  dummies1_JOB<-data.frame(predict(dummyVars(~ JOB, data=infile21),newdata=infile21))
  dummies1_CAR_USE<-data.frame(predict(dummyVars(~ CAR_USE, data=infile21),newdata=infile21))
  dummies1_CAR_TYPE<-data.frame(predict(dummyVars(~ CAR_TYPE, data=infile21),newdata=infile21))
  dummies1_RED_CAR<-data.frame(predict(dummyVars(~ RED_CAR, data=infile21),newdata=infile21))
  dummies1_REVOKED<-data.frame(predict(dummyVars(~ MSTATUS, data=infile21),newdata=infile21))
  dummies1_URBANICITY<-data.frame(predict(dummyVars(~ URBANICITY, data=infile21),newdata=infile21))

     infile31<-cbind(infile21,dummies1_SEX,dummies1_PARENT1,dummies1_MSTATUS,dummies1_EDUCATION,dummies


  infile41<-infile31[ ,c(4,5,7,8,10,17,25,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,4


  preProc <- preProcess(method="bagImpute",infile41)

  infile61 <- predict(preProc, infile41)

   infile21$YOJ<-infile61$YOJ
   infile21$INCOME<-infile61$INCOME
   infile21$HOME_VAL<-infile61$HOME_VAL
   infile21$CAR_AGE<-infile61$CAR_AGE
   infile21$AGE<-infile61$AGE

  infile51<-infile21

  infile51$PARENT1<-as.factor(infile51$PARENT1)
  infile51$MSTATUS<-as.factor(infile51$MSTATUS)
  infile51$SEX<-as.factor(infile51$SEX)
  infile51$JOB<-as.factor(infile51$JOB)
  infile51$CAR_USE<-as.factor(infile51$CAR_USE)
  infile51$CAR_TYPE<-as.factor(infile51$CAR_TYPE)
  infile51$EDUCATION<-as.factor(infile51$EDUCATION)
  infile51$RED_CAR<-as.factor(infile51$RED_CAR)
  infile51$REVOKED<-as.factor(infile51$REVOKED)
  infile51$URBANICITY<-as.factor(infile51$URBANICITY)
  infile51<-cbind(infile51,infile61[,c(18:25)]  )
  infile51<-infile51[,c(-14)]
  str(infile51)
```

```
## 'data.frame':    2141 obs. of  33 variables:
##  $ INDEX          : int  3 9 10 18 21 30 31 37 39 47 ...
##  $ TARGET_FLAG    : logi  NA NA NA NA NA NA ...
##  $ TARGET_AMT     : logi  NA NA NA NA NA NA ...
##  $ KIDSDRIV       : int  0 1 0 0 0 0 0 0 2 0 ...
##  $ AGE            : num  48 40 44 35 59 46 60 54 36 50 ...
```

```
##  $ HOMEKIDS        : int  0 1 2 2 0 0 0 0 2 0 ...
##  $ YOJ             : num  11 11 12 10.9 12 ...
##  $ INCOME          : num  52881 50815 43486 21204 87460 ...
##  $ PARENT1         : Factor w/ 2 levels "No","Yes": 1 2 2 2 1 1 1 1 2 1 ...
##  $ HOME_VAL        : num  0 0 0 0 0 ...
##  $ MSTATUS         : Factor w/ 2 levels "Yes","z_No": 2 2 2 2 2 1 1 1 2 2 ...
##  $ SEX             : Factor w/ 2 levels "M","z_F": 1 1 2 1 1 1 2 1 2 2 ...
##  $ EDUCATION       : Factor w/ 5 levels "<High School",..: 2 5 5 5 5 2 5 1 2 4 ...
##  $ TRAVTIME        : int  26 21 30 74 45 7 16 27 5 22 ...
##  $ CAR_USE         : Factor w/ 2 levels "Commercial","Private": 2 2 1 2 2 1 1 1 1 2 ...
##  $ BLUEBOOK        : num  21970 18930 5900 9230 15420 ...
##  $ TIF             : int  1 6 10 6 1 1 1 4 4 4 ...
##  $ CAR_TYPE        : Factor w/ 6 levels "Minivan","Panel Truck",..: 5 1 6 3 1 2 4 2 1 4 ...
##  $ RED_CAR         : Factor w/ 2 levels "no","yes": 2 1 1 1 2 1 1 1 1 1 ...
##  $ OLDCLAIM        : num  0 3295 0 0 44857 ...
##  $ CLM_FREQ        : int  0 1 0 0 2 1 0 0 0 0 ...
##  $ REVOKED         : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 1 1 1 1 1 ...
##  $ MVR_PTS         : int  2 2 0 0 4 2 0 5 0 3 ...
##  $ CAR_AGE         : num  10 1 10 4 1 ...
##  $ URBANICITY      : Factor w/ 2 levels "Highly Urban/ Urban",..: 1 1 2 2 1 1 1 1 2 1 ...
##  $ JOBClerical     : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ JOBDoctor       : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ JOBHome.Maker   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ JOBLawyer       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ JOBManager      : num  1 1 0 0 1 0 0 0 0 0 ...
##  $ JOBProfessional : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ JOBStudent      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ JOBz_Blue.Collar: num  0 0 1 0 0 0 1 1 1 0 ...
```

```
   infile51$TARGET_FLAG <- predict(fit, newdata=infile51, type="response")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type =
## ifelse(type == : prediction from a rank-deficient fit may be misleading
```

```
   infile51$TARGET_AMT <- predict(fit1,infile51)
```

```
## Warning in predict.lm(fit1, infile51): prediction from a rank-deficient
## fit may be misleading
```

```
   outfile<-infile51[,c(1,2,3)]
   write.csv(outfile, file = "Auto_Insurance.csv",row.names=FALSE)
```

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.