

age.R

narah

2019-09-01

```
##install.packages("xlsx")
##install.packages("sqldf")
##install.packages("imputeTS")
##install.packages("randomForest")
library(randomForest)

## Warning: package 'randomForest' was built under R version 3.6.1
## randomForest 4.6-14

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

library(imputeTS)

## Warning: package 'imputeTS' was built under R version 3.6.1

## Registered S3 methods overwritten by 'ggplot2':
##   method      from
##   [.quosures   rlang
##   c.quosures   rlang
##   print.quosures rlang

## Registered S3 method overwritten by 'xts':
##   method      from
##   as.zoo.xts   zoo

## Registered S3 method overwritten by 'quantmod':
##   method      from
##   as.zoo.data.frame zoo

## Registered S3 methods overwritten by 'forecast':
##   method      from
##   fitted.fracdiff   fracdiff
##   residuals.fracdiff fracdiff

library(sqldf)

## Warning: package 'sqldf' was built under R version 3.6.1
## Loading required package: gsubfn

## Warning: package 'gsubfn' was built under R version 3.6.1
## Loading required package: proto
```

```

## Warning: package 'proto' was built under R version 3.6.1

## Loading required package: RSQLite

## Warning: package 'RSQLite' was built under R version 3.6.1

masterdata_age = read.csv("C:/naraharitransactions/AddmissionsInfo/ALY6080
90571 Integrated Experiential Learn SEC 02 Summer 2019
CPS/Week5/video_age_gender.csv",1)
metadata = read.csv("C:/naraharitransactions/AddmissionsInfo/ALY6080 90571
Integrated Experiential Learn SEC 02 Summer 2019 CPS/Week5/video_meta.csv",1)

##age_data = sqldf("select * from masterdata_age ins join metadata meta
using(hID)")
facebook_age_data = sqldf("select
meta.*,[U.35.44],[U.55.64],[F.25.34],[M.18.24],[M.55.64] from masterdata_age
ins join metadata meta using(hID) where meta.Is_episode = 1")
##View(tvchannel_age_data)

tvchannel_age_data = sqldf("select meta.*,
[U.35.44],[U.55.64],[F.25.34],[M.18.24],[M.55.64] from masterdata_age ins
join metadata meta using(hID) where meta.Is_episode = 0")

facebook_age_data[c(1, 2)] <- list(NULL)
tvchannel_age_data[c(1,2)] <- list(NULL)

## Alternative Hypothesis Ha : muf > mut
t.test(facebook_age_data,tvchannel_age_data,mu = 0,alternate="greater")

##
## Welch Two Sample t-test
##
## data: facebook_age_data and tvchannel_age_data
## t = 12.368, df = 27870, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 1.63614e+17 2.25239e+17
## sample estimates:
## mean of x mean of y
## 6.387865e+16 -1.305479e+17

facebook_age_data = na_replace(facebook_age_data,0)
tvchannel_age_data = na_replace(tvchannel_age_data,0)

set.seed(100)
train <- sample(nrow(facebook_age_data), 0.7*nrow(facebook_age_data), replace
= FALSE)
TrainSet <- facebook_age_data[train,]
ValidSet <- facebook_age_data[-train,]
#View(TrainSet)

```

```

model1 = randomForest(U.35.44 ~ ., data=TrainSet, importance = TRUE)
model1

##
## Call:
## randomForest(formula = U.35.44 ~ ., data = TrainSet, importance = TRUE)
##              Type of random forest: regression
##              Number of trees: 500
## No. of variables tried at each split: 2
##
##              Mean of squared residuals: 3.690016e+15
##              % Var explained: 87.62

model2 = randomForest(U.35.44 ~ ., data=TrainSet, nTree = 500, mTry = 6 ,
importance = TRUE)
model2

##
## Call:
## randomForest(formula = U.35.44 ~ ., data = TrainSet, nTree = 500,
mTry = 6, importance = TRUE)
##              Type of random forest: regression
##              Number of trees: 500
## No. of variables tried at each split: 2
##
##              Mean of squared residuals: 3.795456e+15
##              % Var explained: 87.27

predictTrainSet = predict(model1,TrainSet,Type = "class")

##table(predictTrainSet, TrainSet$U.35.44)
prdedictvalidSet = predict(model2,ValidSet, Type = "class")

#table(prdedictvalidSet, ValidSet$total_video_views_unique)

## Predicting with TV Channels Data

PredictTvChannelsData = predict(model2,tvchannel_age_data,Type = "class")
##PredictTvChannelsData

```