Project.R

narah

2019-09-01

```r
##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.
```

```r
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
```

```r
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_ins  = read.csv("C:/naraharitransactions/AddmissionsInfo/ALY6080 9
0571 Integrated Experiential Learn SEC 02 Summer 2019 CPS/Week5/video_insight
s.csv",1)
metadata = read.csv("C:/naraharitransactions/AddmissionsInfo/ALY6080 90571 In
tegrated Experiential Learn SEC 02 Summer 2019 CPS/Week5/video_meta.csv",1)

insights_data =  sqldf("select meta.*, ins.* from masterdata_ins ins  join me
tadata  meta using(hID)")
##write.csv2(insights_data, "C:/naraharitransactions/AddmissionsInfo/ALY6080
90571 Integrated Experiential Learn SEC 02 Summer 2019 CPS/Week5/combined_ins
ight.csv")
facebook_insights_data =  sqldf("select meta.*, total_video_views_unique,ins.
total_video_views_autoplayed,ins.total_video_views,ins.total_video_views_paid
,total_video_views_sound_on,total_video_play_count,[total_video_stories_by_ac
tion_type.like],[total_video_reactions_by_type_total.like] from masterdata_in
s ins  join metadata  meta using(hID) where meta.Is_episode = 1")
##write.csv2(facebook_insights_data, "C:/naraharitransactions/AddmissionsInfo
/ALY6080 90571 Integrated Experiential Learn SEC 02 Summer 2019 CPS/Week5/fac
ebook_insight.csv")

tvchannel_insights_data =  sqldf("select meta.*, total_video_views_unique, in
s.total_video_views_autoplayed,ins.total_video_views,ins.total_video_views_pa
id,total_video_views_sound_on,total_video_play_count,[total_video_stories_by_
action_type.like],[total_video_reactions_by_type_total.like]  from masterdata
_ins ins  join metadata  meta using(hID) where meta.Is_episode = 0")
##write.csv2(tvchannel_insights_data, "C:/naraharitransactions/AddmissionsInf
o/ALY6080 90571 Integrated Experiential Learn SEC 02 Summer 2019 CPS/Week5/tv
channel_insights_data.csv")
##str(tvchannel_insights_data)
##View(facebook_insights_data)
facebook_insights_data[c(1, 2)] <- list(NULL)
tvchannel_insights_data[c(1,2)] <- list(NULL)

## Alternative Hypothesis Ha : muf > mut
t.test(facebook_insights_data,tvchannel_insights_data,mu = 0,alternate="great
er")

##
##   Welch Two Sample t-test
##
## data:  facebook_insights_data and tvchannel_insights_data
## t = 12.468, df = 37381, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##   1.241100e+17 1.704097e+17
## sample estimates:
##     mean of x     mean of y
##  4.791535e+16 -9.934452e+16

##facebook_insights_data
##test = rbind (facebook_insights_data, colSums(facebook_insights_data[1:3,])
,colSums(facebook_insights_data))
##columnTOtal = rowsum(facebook_insights_data)
#facebook_insights_data$total <- with(facebook_insights_data, rowSums(faceboo
k_insights_data))
##facebook_insights_data =merge(facebook_insights_data,rowsum(facebook_insigh
ts_data)) ##, by="date_updated")
##View(facebook_insights_data$total)
##View(columnTOtal)




insights_data <- na_replace(insights_data,0) # replace with ""
facebook_insights_data = na_replace(facebook_insights_data,0)
tvchannel_insights_data = na_replace(tvchannel_insights_data,0)

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

model1  =  randomForest(total_video_views_unique ~ ., data=TrainSet, importan
ce = TRUE)
model1

##
## Call:
##  randomForest(formula = total_video_views_unique ~ ., data = TrainSet,
importance = TRUE)
##               Type of random forest: regression
##                     Number of trees: 500
## No. of variables tried at each split: 3
##
##           Mean of squared residuals: 1.640163e+12
##                   % Var explained: 91.53

model2  = randomForest(total_video_views_unique ~ ., data=TrainSet, nTree = 5
00, mTry = 6 , importance = TRUE)
model2

##
## Call:
```

```
##  randomForest(formula = total_video_views_unique ~ ., data = TrainSet,
nTree = 500, mTry = 6, importance = TRUE)
##                  Type of random forest: regression
##                        Number of trees: 500
## No. of variables tried at each split: 3
##
##            Mean of squared residuals: 1.975511e+12
##                       % Var explained: 89.8

predictTrainSet = predict(model2,TrainSet,Type = "class")

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

#table(prdedictvalidSet, ValidSet$total_video_views_unique)

## Predicting with TV Channels Data

PredictTvChannelsData =  predict(model2,tvchannel_insights_data,Type = "class
")
##table(PredictTvChannelsData, tvchannel_insights_data$total_video_views_uniq
ue)
#PredictTvChannelsData
```