|  |
| --- |
|  |
| **Assignment 7.3** |  |
|  |  |
|  | library(dplyr); library(corrplot);library(car); library(MASS) |
|  | # Q1- read the dataset and identify the right features |
|  |  |
|  | # import train data set |
|  | Variant\_1 <- read.csv("E:/Data Analytics with RET/Assignment/Dataset/fbtrain/Features\_Variant\_1.csv", header=FALSE) |
|  | Variant\_2 <- read.csv("E:/Data Analytics with RET/Assignment/Dataset/fbtrain/Features\_Variant\_2.csv", header=FALSE) |
|  | Variant\_3 <- read.csv("E:/Data Analytics with RET/Assignment/Dataset/fbtrain/Features\_Variant\_3.csv", header=FALSE) |
|  | Variant\_4 <- read.csv("E:/Data Analytics with RET/Assignment/Dataset/fbtrain/Features\_Variant\_4.csv", header=FALSE) |
|  | Variant\_5 <- read.csv("E:/Data Analytics with RET/Assignment/Dataset/fbtrain/Features\_Variant\_5.csv", header=FALSE) |
|  | fbtrain <- rbind(Variant\_1, Variant\_2, Variant\_3, Variant\_4, Variant\_5) |
|  | dim(fbtrain) |
|  |  |
|  | # import test data set |
|  | setwd("E:/Data Analytics with RET/Assignment/Dataset/fbtest") |
|  | test1 <- read.csv("Test\_Case\_1.csv", header = F); test2 <- read.csv("Test\_Case\_2.csv", header = F) |
|  | test3 <- read.csv("Test\_Case\_3.csv", header = F); test4 <- read.csv("Test\_Case\_4.csv", header = F) |
|  | test5 <- read.csv("Test\_Case\_5.csv", header = F); test6 <- read.csv("Test\_Case\_6.csv", header = F) |
|  | test7 <- read.csv("Test\_Case\_7.csv", header = F); test8 <- read.csv("Test\_Case\_8.csv", header = F) |
|  | test9 <- read.csv("Test\_Case\_9.csv", header = F); test10 <- read.csv("Test\_Case\_10.csv", header = F) |
|  | fbtest <- rbind(test1, test2, test3, test4, test5, test6, test7, test8, test9, test10) |
|  | dim(fbtest) |
|  |  |
|  | # Assign variable names to the train and test data set |
|  | colnames(fbtrain) <- c("plikes","checkin","talking","category","d5","d6","d7","d8","d9","d10","d11","d12", |
|  | "d13","d14","d15","d16","d17","d18","d19","d20","d21","d22","d23","d24","d25","d26", |
|  | "d27","d28","d29","cc1","cc2","cc3","cc4","cc5","basetime","postlength","postshre", |
|  | "postpromo","Hhrs","sun","mon","tue","wed","thu","fri","sat","basesun","basemon", |
|  | "basetue","basewed","basethu","basefri","basesat","target") |
|  | colnames(fbtest) <- c("plikes","checkin","talking","category","d5","d6","d7","d8","d9","d10","d11","d12", |
|  | "d13","d14","d15","d16","d17","d18","d19","d20","d21","d22","d23","d24","d25","d26", |
|  | "d27","d28","d29","cc1","cc2","cc3","cc4","cc5","basetime","postlength","postshre", |
|  | "postpromo","Hhrs","sun","mon","tue","wed","thu","fri","sat","basesun","basemon", |
|  | "basetue","basewed","basethu","basefri","basesat","target") |
|  |  |
|  | dim(fbtrain); dim(fbtest) |
|  | View(fbtrain); View(fbtest) |
|  | str(fbtrain); str(fbtest) |
|  |  |
|  | train <- fbtrain; test <- fbtest |
|  | head(train); head(test) |
|  |  |
|  | # making the data tidy by constructing single collumn for post publish day |
|  | train$pubday<- ifelse(train$sun ==1, 1, ifelse(train$mon ==1, 2, ifelse(train$tue ==1, 3, |
|  | ifelse(train$wed ==1, 4, ifelse(train$thu ==1, 5, ifelse(train$fri ==1, 6, |
|  | ifelse(train$sat ==1, 7, NA))))))) |
|  | # making the data tidy by constructing single collumn for base day |
|  | train$baseday<- ifelse(train$basesun ==1, 1, ifelse(train$basemon ==1, 2, ifelse(train$basetue ==1, 3, |
|  | ifelse(train$basewed ==1, 4, ifelse(train$basethu ==1, 5, |
|  | ifelse(train$basefri ==1, 6, ifelse(train$basesat ==1, 7, NA))))))) |
|  | # now the data set is ready |
|  | #-------------------------------------------------------------------- |