



ACADGILD

SESSION 13: Decision Tree Based Models

Assignment 2

Submitted by: Udaykumar

Udaykumarr019@gmail.com

(M):+91-8123431864

Data Analytics

Table of Contents

| | |
|----------------------------|---|
| 1. Problem Statement | 3 |
| 2. Solution | 3 |

1. Problem Statement

Use the given link below:

<https://archive.ics.uci.edu/ml/machine-learning-databases/00304/>

Problem- prediction of the number of comments in the upcoming 24 hours on those blogs, the train data was generated from different base times that may temporally overlap. Therefore, if you simply split the train into disjoint partitions, the underlying time intervals may overlap. Therefore, they should use the provided, temporally disjoint train and test splits to ensure that the evaluation is fair.

- a) Create a linear regression model to predict the number of comments in the next 24 hours (relative to base time).
- b) Fine tune the model and represent important features Visualize the dataset and make inferences from that.
- c) Interpret the summary of the linear model.
- d) Report the test accuracy vs. the training accuracy

2. Solution

- a) **Create a linear regression model to predict the number of comments in the next 24 hours (relative to base time).**

The R-script for the given problem is as follows:

```
library(data.table)
library(foreach)
library(readr)
library(dplyr)
library(corrplot); library(car); library(MASS); library(ggplot2)
library(reshape2); library(forecast)

setwd("E:/uday/acadgild data analytics/supporting files/BlogFeedback") getwd()

blogData_train <- read_csv("E:/uday/acadgild data analytics/supporting
files/BlogFeedback/blogData_train.csv")
```

```
# retrieve filenames of test sets
```

```
test_filenames = list.files(pattern = "blogData_test")
```

```
# load and combine dataset
```

```
train = fread("blogData_train.csv")
```

```
fbtest = foreach(i = 1:length(test_filenames), .combine = rbind) %do% {
```

```
  temp = fread(test_filenames[i], header = FALSE)
```

```
}
```

```
# Assign variable names to the train and test data set
```

```
colnames(blogData_train) <-
```

```
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", "V55", "V56", "V57", "V58",
```

```
"V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58",
```

```
"V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55",  
", "V56",
```

```
"V57", "V58", "V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55", "V56", "V57",  
", "V58",
```

```
"V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55",  
"V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55", "V56", "V57",  
"V58", "V55", "V56", "V57", "V58")
```

```
dim(blogData_train)
```

```
dim(fbtest)
```

```
View(blogData_train)
```

```
View(fbtest)
```

```
str(blogData_train)
```

```

str(fbtest)

train <- blogData_train; test <- fbtest
head(train); head(test)

# making the data tidy by constructing single column 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 column 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))))))

# a. Create a linear regression model to predict the number of comments in the next 24
hours
# (relative to basetime)
#install.packages(MASS)
library(MASS)

final_model <- lm(target ~ checkin + talking + d5 + d6 + d7 + d8 + d9 + d10 + d11 +
                  d12 + d13 + d16 + d17 + d19 + d20 + d21 + d22 + d23 + d24 +
                  cc1 + cc2 + cc3 + cc4 + basetime + postshre + Hhrs + wed +
                  thu + fri + basemon + basewed, data = train)
summary(final_model)

```

The output of the R-Script (from Console window) is given as follows:

```

> library(data.table)
> library(foreach)
> library(readr)
> library(dplyr)
> library(corrplot); library(car); library(MASS); library(ggplot2)
> library(reshape2); library(forecast)
>
> setwd("E:/uday/acadgild data analytics/supporting files/BlogFeedback")
> getwd()
[1] "E:/uday/acadgild data analytics/supporting files/BlogFeedback"
>
> blogData_train <- read_csv("E:/uday/acadgild data analytics/supporting
files/BlogFeedback/blogData_train.csv") Parsed with column specification:
cols(
  .default = col_double()
)
See spec(...) for full column specifications.
=====
=====| 100%
62 MB
>

```

```
> # retrieve filenames of test sets
> test_filenames = list.files(pattern = "blogData_test")
>
> # load and combine dataset
> train = fread("blogData_train.csv")
> fbtest = foreach(i = 1:length(test_filenames), .combine = rbind) %do% {
+   temp = fread(test_filenames[i], header = FALSE)
+ }
>
>
> # Assign variable names to the train and test data set
> colnames(blogData_train) <-
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", "post shre",
+ "postpromo", "Hhrs", "sun", "mon", "tue", "wed", "thu", "fri", "sat", "basesun", "basem on",
+ "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", "post shre",
+ "postpromo", "Hhrs", "sun", "mon", "tue", "wed", "thu", "fri", "sat", "basesun", "basem on",
+ "basetue", "basewed", "basethu", "basefri", "basesat", "target", "V55", "V56", "V57", "V58",
+ "V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58",
+ "V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55",
+ "V56",
+ "V57", "V58", "V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55", "V56", "V57",
+ "V58",
+ "V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55", "V56", "V57", "V58", "V55",
+ "V56", "V58", "V55", "V56", "V57", "V58")
>
> dim(blogData_train) [1]
52396      281
> dim(fbtest)
[1] 7624      281
> View(blogData_train)
```

| | plikes | checkin | talking | category | d5 | d6 | d7 | d8 | d9 | d10 | d11 | d12 | d13 | d14 | d15 | d16 | d17 | d18 | d19 | d20 | d21 | d22 |
|----|----------|----------|---------|----------|----|----------|----------|----|-----|-----|----------|----------|-----|-----|-----|----------|----------|-----|-----|-----|----------|----------|
| 1 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 2 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 3 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 4 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 5 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 6 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 7 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 8 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 9 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 10 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 11 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 12 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 13 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 14 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 15 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 16 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 17 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 18 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 19 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 20 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 21 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |
| 22 | 40.30467 | 53.84566 | 0 | 401 | 15 | 15.52416 | 32.44188 | 0 | 377 | 3 | 14.04423 | 32.61542 | 0 | 377 | 2 | 34.56757 | 48.47518 | 0 | 378 | 12 | 1.479934 | 46.18691 |

Showing 1 to 23 of 52,396 entries

> View(fbtest)

Unlited1*blogData_trainfbtestAssignment 13.2.RpredictedUnlited3*

Filter

| | plikes | checkin | talking | category | d5 | d6 | d7 | d8 | d9 | d10 | d11 | d12 | d13 | d14 | d15 | d16 | d17 | d18 | d19 | d20 | d21 | d22 |
|----|--------------|---------|---------|-------------|----|----|----|------|------|-------------|-------------|-----|-----|-----|------|-----|-----|------|-------------|-----|-----|-----|
| 1 | 10.63066000 | 0 | 0 | 17.8829920 | 1 | 0 | 0 | 259 | 5.0 | 4.01827600 | 10.3967900 | 0 | 0 | 0 | 235 | 0 | 0 | 1.0 | 3.81723950 | 0 | 0 | |
| 2 | 43.43582500 | 0 | 0 | 75.5904850 | 0 | 0 | 0 | 634 | 20.0 | 15.99858950 | 44.5608700 | 0 | 0 | 0 | 473 | 0 | 0 | 2.0 | 15.46967600 | 0 | 1 | |
| 3 | 1.73333330 | 0 | 0 | 3.0433900 | 0 | 0 | 1 | 9 | 0.0 | 0.73333335 | 1.5260698 | 0 | 0 | 0 | 5 | 0 | 0 | 0.0 | 0.66666670 | 0 | 1 | |
| 4 | 27.23021500 | 0 | 0 | 45.9709500 | 0 | 0 | 1 | 371 | 14.0 | 10.78417300 | 24.2099420 | 0 | 0 | 0 | 228 | 0 | 0 | 4.0 | 9.99760150 | 0 | 0 | |
| 5 | 4.50000000 | 0 | 0 | 6.6770754 | 0 | 0 | 1 | 18 | 0.5 | 3.00000000 | 4.0000000 | 0 | 0 | 0 | 10 | 0 | 0 | 0.5 | 1.33333340 | 0 | 0 | |
| 6 | 156.40298000 | 0 | 0 | 246.0559800 | 0 | 0 | 1 | 970 | 28.0 | 76.14925400 | 131.9008300 | 0 | 1 | 0 | 725 | 0 | 0 | 16.0 | 53.32835800 | 0 | 1 | |
| 7 | 10.50931600 | 0 | 0 | 36.5939830 | 0 | 0 | 1 | 191 | 1.0 | 3.60248450 | 20.6338310 | 0 | 0 | 0 | 179 | 0 | 0 | 0.0 | 3.68944100 | 0 | 1 | |
| 8 | 123.86919000 | 0 | 0 | 129.5662200 | 0 | 0 | 1 | 1065 | 87.0 | 43.32897000 | 62.7741470 | 0 | 0 | 0 | 491 | 0 | 0 | 19.5 | 41.18556200 | 0 | 0 | |
| 9 | 22.46341500 | 0 | 0 | 42.1849000 | 0 | 0 | 0 | 188 | 7.5 | 8.21951200 | 25.0204930 | 0 | 0 | 0 | 174 | 0 | 0 | 1.5 | 7.89024400 | 0 | 0 | |
| 10 | 0.00000000 | 0 | 0 | 0.0000000 | 0 | 0 | 1 | 0 | 0.0 | 0.00000000 | 0.0000000 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.00000000 | 0 | 1 | |
| 11 | 0.15550756 | 0 | 0 | 0.6683261 | 0 | 0 | 0 | 7 | 0.0 | 0.07559396 | 0.4113776 | 0 | 0 | 0 | 5 | 0 | 0 | 0.0 | 0.06047516 | 0 | 0 | |
| 12 | 16.59357500 | 0 | 0 | 19.6713640 | 1 | 0 | 0 | 144 | 10.0 | 6.51244970 | 11.0512150 | 0 | 0 | 0 | 111 | 0 | 0 | 2.0 | 5.82570270 | 0 | 0 | |
| 13 | 0.37869823 | 0 | 0 | 1.0817565 | 0 | 0 | 1 | 4 | 0.0 | 0.03550296 | 0.2146551 | 0 | 0 | 0 | 2 | 0 | 0 | 0.0 | 0.03550296 | 0 | 0 | |
| 14 | 49.44236800 | 0 | 0 | 112.6201250 | 1 | 0 | 0 | 849 | 9.0 | 20.44548200 | 62.6193900 | 0 | 0 | 0 | 506 | 0 | 0 | 2.0 | 19.22118400 | 0 | 0 | |
| 15 | 122.81293000 | 0 | 0 | 109.9611000 | 0 | 0 | 1 | 1069 | 89.0 | 44.89454300 | 74.5475300 | 0 | 0 | 0 | 1046 | 0 | 0 | 12.0 | 42.81889000 | 0 | 0 | |
| 16 | 56.51209300 | 0 | 0 | 77.4428300 | 0 | 0 | 1 | 438 | 32.0 | 19.29653000 | 49.2213440 | 0 | 0 | 0 | 432 | 0 | 0 | 0.0 | 18.96635000 | 0 | 0 | |
| 17 | 43.43582500 | 0 | 0 | 75.5904850 | 0 | 0 | 1 | 634 | 20.0 | 15.99858950 | 44.5608700 | 0 | 0 | 0 | 473 | 0 | 0 | 2.0 | 15.46967600 | 0 | 1 | |
| 18 | 10.63066000 | 0 | 0 | 17.8829920 | 1 | 0 | 0 | 259 | 5.0 | 4.01827600 | 10.3967900 | 0 | 0 | 0 | 235 | 0 | 0 | 1.0 | 3.81723950 | 0 | 0 | |
| 19 | 122.81293000 | 0 | 0 | 109.9611000 | 0 | 0 | 1 | 1069 | 89.0 | 44.89454300 | 74.5475300 | 0 | 0 | 0 | 1046 | 0 | 0 | 12.0 | 42.81889000 | 0 | 0 | |
| 20 | 122.81293000 | 0 | 0 | 109.9611000 | 0 | 0 | 1 | 1069 | 89.0 | 44.89454300 | 74.5475300 | 0 | 0 | 0 | 1046 | 0 | 0 | 12.0 | 42.81889000 | 0 | 0 | |
| 21 | 0.00000000 | 0 | 0 | 0.0000000 | 0 | 0 | 1 | 0 | 0.0 | 0.00000000 | 0.0000000 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0 | 0.00000000 | 0 | 0 | |
| 22 | 0.37869823 | 0 | 0 | 1.0817565 | 0 | 0 | 1 | 4 | 0.0 | 0.03550296 | 0.2146551 | 0 | 0 | 0 | 2 | 0 | 0 | 0.0 | 0.03550296 | 0 | 0 | |

Showing 1 to 23 of 7,624 entries

Activate Windows

Showing 1 to 23 of 7,624 entries

> str(blogData_train)

Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame':

52396 obs. of

281 variables:

```

$ plikes      : num  40.3 40.3 40.3 40.3 40.3 ...
$ checkin    : num  53.8 53.8 53.8 53.8 53.8 ...
$ talking     : num  0 0 0 0 0 0 0 0 0 0 ...
$ category    : num  401 401 401 401 401 401 401 401 401 401 ...
$ d5          : num  15 15 15 15 15 15 15 15 15 15 ...
$ d6          : num  15.5 15.5 15.5 15.5 15.5 15.5 15.5 15.5 15.5 15.5 ...
$ d7          : num  32.4 32.4 32.4 32.4 32.4 32.4 32.4 32.4 32.4 32.4 ...
$ d8          : num  0 0 0 0 0 0 0 0 0 0 ...
$ d9          : num  377 377 377 377 377 377 377 377 377 377 ...
$ d10         : num  3 3 3 3 3 3 3 3 3 3 ...
$ d11         : num  14 14 14 14 14 14 14 14 14 14 ...
$ d12         : num  32.6 32.6 32.6 32.6 32.6 32.6 32.6 32.6 32.6 32.6 ...
$ d13         : num  0 0 0 0 0 0 0 0 0 0 ...
$ d14         : num  377 377 377 377 377 377 377 377 377 377 ...
$ d15         : num  2 2 2 2 2 2 2 2 2 2 ...
$ d16         : num  34.6 34.6 34.6 34.6 34.6 34.6 34.6 34.6 34.6 34.6 ...
$ d17         : num  48.5 48.5 48.5 48.5 48.5 48.5 48.5 48.5 48.5 48.5 ...
$ d18         : num  0 0 0 0 0 0 0 0 0 0 ...
$ d19         : num  378 378 378 378 378 378 378 378 378 378 ...

```


\$ NA : num 0 0 0 0 0 0 0 0 0 0 0 ...

```

$ NA      : num      0 0 0 0 0 0 0 0 0 0 ...
$ NA      : num      0 0 0 0 0 0 0 0 0 0 ...
$ NA      : num      0 0 0 0 0 0 0 0 0 0 ...
$ NA      : num      0 0 0 0 0 0 0 0 0 0 ...
$ NA      : num      0 0 0 0 0 0 0 0 0 0 ...
$ NA      : num      0 0 0 0 0 0 0 0 0 0 ...
$ NA      : num      0 0 0 0 0 0 0 0 0 0 ...
$ NA      : num      0 0 0 0 0 0 0 0 0 0 ...

```

```
[list output truncated]
```

```
- attr(*,"spec")=
```

```

.. cols(
..   `40.30467` = col_double(),
..   `53.845657` = col_double(),
..   `0.0` = col_double(),
..   `401.0` = col_double(),
..   `15.0` = col_double(),
..   `15.52416` = col_double(),
..   `32.44188` = col_double(),
..   `0.0_1` = col_double(),
..   `377.0` = col_double(),
..   `3.0` = col_double(),
..   `14.044226` = col_double(),
..   `32.615417` = col_double(),
..   `0.0_2` = col_double(),
..   `377.0_1` = col_double(),
..   `2.0` = col_double(),
..   `34.567566` = col_double(),
..   `48.475178` = col_double(),
..   `0.0_3` = col_double(),
..   `378.0` = col_double(),
..   `12.0` = col_double(),
..   `1.4799345` = col_double(),
..   `46.18691` = col_double(),
..   `-356.0` = col_double(),
..   `377.0_2` = col_double(),
..   `0.0_4` = col_double(),
..   `1.0761671` = col_double(),
..   `1.795416` = col_double(),
..   `0.0_5` = col_double(),
..   `11.0` = col_double(),
..   `0.0_6` = col_double(),
..   `0.4004914` = col_double(),
..   `1.0780969` = col_double(),
..   `0.0_7` = col_double(),
..   `9.0` = col_double(),
..   `0.0_8` = col_double(),
..   `0.37755936` = col_double(),
..   `1.07421` = col_double(),
..   `0.0_9` = col_double(),
..   `9.0_1` = col_double(),
..   `0.0_10` = col_double(),
..   `0.972973` = col_double(),
..   `1.704671` = col_double(),
..   `0.0_11` = col_double(),
..   `10.0` = col_double(),
..   `0.0_12` = col_double(),
..   `0.022932023` = col_double(),
..   `1.521174` = col_double(),
..   `-8.0` = col_double(),
..   `9.0_2` = col_double(),
..   `0.0_13` = col_double(),
..   `2.0_1` = col_double(),
..   `2.0_2` = col_double(),
..   `0.0_14` = col_double(),
..   `2.0_3` = col_double(),
..   `2.0_4` = col_double(),
..   `0.0_15` = col_double(),
..   `0.0_16` = col_double(),
..   `0.0_17` = col_double(),
..   `0.0_18` = col_double(),
..   `0.0_19` = col_double(),
..   `10.0_1` = col_double(),

```

```

..      0.0_20 = col_double(),
..      0.0_21 = col_double(),
..      0.0_22 = col_double(),
..      0.0_23 = col_double(),
..      0.0_24 = col_double(),
..      0.0_25 = col_double(),
..      0.0_26 = col_double(),
..      0.0_27 = col_double(),
..      0.0_28 = col_double(),
..      0.0_29 = col_double(),
..      0.0_30 = col_double(),
..      0.0_31 = col_double(),
..      0.0_32 = col_double(),
..      0.0_33 = col_double(),
..      0.0_34 = col_double(),
..      0.0_35 = col_double(),
..      0.0_36 = col_double(),
..      0.0_37 = col_double(),
..      0.0_38 = col_double(),
..      0.0_39 = col_double(),
..      0.0_40 = col_double(),
..      0.0_41 = col_double(),
..      0.0_42 = col_double(),
..      0.0_43 = col_double(),
..      0.0_44 = col_double(),
..      0.0_45 = col_double(),
..      0.0_46 = col_double(),
..      0.0_47 = col_double(),
..      0.0_48 = col_double(),
..      0.0_49 = col_double(),
..      0.0_50 = col_double(),
..      0.0_51 = col_double(),
..      0.0_52 = col_double(),
..      0.0_53 = col_double(),
..      0.0_54 = col_double(),
..      0.0_55 = col_double(),
..      0.0_56 = col_double(),
..      0.0_57 = col_double(),
..      0.0_58 = col_double(),
..      0.0_59 = col_double(),
..      0.0_60 = col_double(),
..      0.0_61 = col_double(),
..      0.0_62 = col_double(),
..      0.0_63 = col_double(),
..      0.0_64 = col_double(),
..      0.0_65 = col_double(),
..      0.0_66 = col_double(),
..      0.0_67 = col_double(),
..      0.0_68 = col_double(),
..      0.0_69 = col_double(),
..      0.0_70 = col_double(),
..      0.0_71 = col_double(),
..      0.0_72 = col_double(),
..      0.0_73 = col_double(),
..      0.0_74 = col_double(),
..      0.0_75 = col_double(),
..      0.0_76 = col_double(),
..      0.0_77 = col_double(),
..      0.0_78 = col_double(),
..      0.0_79 = col_double(),
..      0.0_80 = col_double(),
..      0.0_81 = col_double(),
..      0.0_82 = col_double(),
..      0.0_83 = col_double(),
..      0.0_84 = col_double(),
..      0.0_85 = col_double(),
..      0.0_86 = col_double(),
..      0.0_87 = col_double(),
..      0.0_88 = col_double(),
..      0.0_89 = col_double(),
..      0.0_90 = col_double(),
..      0.0_91 = col_double(),

```



```

..      0.0_164` = col_double(),
..      0.0_165` = col_double(),
..      0.0_166` = col_double(),
..      0.0_167` = col_double(),
..      0.0_168` = col_double(),
..      0.0_169` = col_double(),
..      0.0_170` = col_double(),
..      0.0_171` = col_double(),
..      0.0_172` = col_double(),
..      0.0_173` = col_double(),
..      0.0_174` = col_double(),
..      0.0_175` = col_double(),
..      0.0_176` = col_double(),
..      0.0_177` = col_double(),
..      0.0_178` = col_double(),
..      0.0_179` = col_double(),
..      0.0_180` = col_double(),
..      0.0_181` = col_double(),
..      0.0_182` = col_double(),
..      0.0_183` = col_double(),
..      0.0_184` = col_double(),
..      0.0_185` = col_double(),
..      0.0_186` = col_double(),
..      0.0_187` = col_double(),
..      0.0_188` = col_double(),
..      0.0_189` = col_double(),
..      0.0_190` = col_double(),
..      0.0_191` = col_double(),
..      0.0_192` = col_double(),
..      0.0_193` = col_double(),
..      0.0_194` = col_double(),
..      0.0_195` = col_double(),
..      0.0_196` = col_double(),
..      0.0_197` = col_double(),
..      0.0_198` = col_double(),
..      0.0_199` = col_double(),
..      0.0_200` = col_double(),
..      0.0_201` = col_double(),
..      0.0_202` = col_double(),
..      0.0_203` = col_double(),
..      0.0_204` = col_double(),
..      0.0_205` = col_double(),
..      0.0_206` = col_double(),
..      0.0_207` = col_double(),
..      0.0_208` = col_double(),
..      0.0_209` = col_double(),
..      0.0_210` = col_double(),
..      0.0_211` = col_double(),
..      0.0_212` = col_double(),
..      0.0_213` = col_double(),
..      0.0_214` = col_double(),
..      0.0_215` = col_double(),
..      0.0_216` = col_double(),
..      0.0_217` = col_double(),
..      0.0_218` = col_double(),
..      0.0_219` = col_double(),
..      0.0_220` = col_double(),
..      0.0_221` = col_double(),
..      0.0_222` = col_double(),
..      0.0_223` = col_double(),
..      0.0_224` = col_double(),
..      `1.0` = col_double(),
..      `0.0_225` = col_double(),
..      `0.0_226` = col_double(),
..      `0.0_227` = col_double(),
..      `0.0_228` = col_double(),
..      `0.0_229` = col_double(),
..      `1.0_1` = col_double(),
..      `0.0_230` = col_double(),
..      `0.0_231` = col_double(),
..      `0.0_232` = col_double(),
..      `0.0_233` = col_double(),

```

```

..      `0.0_234` = col_double(),
..      `0.0_235` = col_double(),
..      `0.0_236` = col_double(),
..      `1.0_2` = col_double()
.. )
> str(fbtest)
Classes 'data.table' and 'data.frame': 7624 obs. of      281 variables:
 $ plikes      : num  10.63 43.44 1.73 27.23 4.5 ...
 $ checkin     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ talking     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ category    : num  17.88 75.59 3.04 45.97 6.68 ...
 $ d5          : num  1 0 0 0 0 0 0 0 0 0 ...
 $ d6          : num  259 634 9 371 18 ...
 $ d7          : num  5 20 0 14 0.5 28 1 87 7.5 0 ...
 $ d8          : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d9          : num  0 0 0 0 0 1 0 0 0 0 ...
 $ d10         : num  4.018 15.999 0.733 10.784 3 ...
 $ d11         : num  10.4 44.56 1.53 24.21 4 ...
 $ d12         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d13         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d14         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d15         : num  235 473 5 228 10 725 179 491 174 0 ...
 $ d16         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d17         : num  0 1 1 0 0 1 1 0 0 1 ...
 $ d18         : num  1 2 0 4 0.5 16 0 19.5 1.5 0 ...
 $ d19         : num  3.817 15.47 0.667 9.998 1.333 ...
 $ d20         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d21         : num  0 0 1 0 0 1 0 0 0 0 ...
 $ d22         : num  10.3 44.69 1.53 24.4 2.56 ...
 $ d23         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d24         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d25         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d26         : num  235 473 5 228 7 725 179 491 174 0 ...
 $ d27         : num  1 1 0 2 0 3 0 14 1 0 ...
 $ d28         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d29         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ cc1         : num  9.78 40.97 1.13 22.56 2.83 ...
 $ cc2         : num  16.07 70.31 1.82 39.76 3.67 ...
 $ cc3         : num  0 0 1 1 0 1 1 0 0 1 ...
 $ cc4         : num  0 0 1 0 0 1 0 0 0 0 ...
 $ cc5         : num  1 0 0 0 0 0 0 0 0 0 ...
 $ basetime    : num  192 479 5 337 8 913 189 786 186 0 ...
 $ postlength  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ postshre    : num  0 0 0 0 0 0 0 0 0 0 ...
 $ postpromo   : num  5 18 0 10 0.5 26 0 74 5.5 0 ...
 $ Hhrs        : num  0.201 0.5289 0.0667 0.7866 1.6667 ...
 $ sun         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ mon         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ tue         : num  13.95 62.13 1.73 30.36 2.21 ...
 $ wed         : num  -229 -461 -5 -156 0 -519 -178 -418 -161 0 ...
 $ thu         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ fri         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ sat         : num  217 473 4 228 6 725 170 491 174 0 ...
 $ basesun     : num  0 0 0 0 0.5 2 0 -3 0 0 ...
 $ basemon     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ basetue     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ basewed     : num  0.252 0.193 0.333 0.11 0 ...
 $ basethu     : num  0.904 0.458 0.699 0.356 0 ...
 $ basefri     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ basesat     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ target      : num  0 0 0 0 0 0 0 0 0 0 ...
 $ V55         : num  14 2 2 2 0 0 6 0 1 0 ...
 $ V56         : num  0 0 1 0 0 1 0 0 0 0 ...
 $ V57         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ V58         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ V55         : num  0.0944 0.0733 0.1333 0.0432 0 ...
 $ V56         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ V57         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ V58         : num  0.507 0.286 0.34 0.215 0 ...
 $ V55         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ V56         : num  0 0 0 0 0 0 0 0 1 0 ...
 $ V57         : num  0 0 0 0 0 0 0 0 0 0 ...

```

```

$ V58      : num 12 2 1 2 0 0 5 0 1 0 ...
$ V55      : num 0 0 0 0 0 0 0 0 0 0 ...
$ V56      : num 0 0 0 0 0 0 0 0 0 0 ...
$ V57      : num 0 0 0 0 0 0 0 0 0 0 ...
$ V58      : num 0.0919 0.0677 0.1333 0.0408 0 ...
$ V55      : num 0.504 0.278 0.34 0.21 0 ...
$ V56      : num 0 0 0 0 0 0 0 0 0 ...
$ V57      : num 0 0 0 0 0 0 0 0 0 ...
$ V58      : num 0 0 0 0 0 0 0 0 0 ...
$ V55      : num 12 2 1 2 0 0 5 0 1 0 ...
$ V56      : num 0 0 1 0 0 1 1 0 0 0 ...
$ V57      : num 0 0 1 0 0 0 0 0 0 0 ...
$ V58      : num 0 0 0 0 0 0 0 0 0 ...
$ V55      : num 0.2335 0.1763 0.2 0.0983 0 ...
$ V56      : num 0 0 0 0 0 0 0 0 0 ...
$ V57      : num 0 0 0 0 0 1 0 0 0 0 ...
$ V58      : num 0.855 0.43 0.4 0.321 0 ...
$ V55      : num 0 0 0 0 0 0 0 0 0 ...
$ V56      : num 0 0 0 0 0 0 0 0 0 ...
$ V57      : num 0 0 0 0 0 0 0 0 0 ...
$ V58      : num 13 2 1 2 0 0 5 0 1 0 ...
$ V55      : num 0 0 0 0 0 0 0 0 0 ...
$ V56      : num 0 0 0 0 0 0 0 0 0 ...
$ V57      : num 0 0 0 0 0 0 0 0 0 ...
$ V58      : num 0.00245 0.00564 0 0.0024 0 ...
$ V55      : num 0.675 0.404 0.365 0.29 0 ...
$ V56      : num 0 0 0 0 0 0 0 0 0 ...
$ V57      : num 0 0 0 0 0 0 0 0 0 ...
$ V58      : num -10 -2 -1 -2 0 0 -5 0 -1 0 ...
$ V55      : num 12 2 1 2 0 0 5 0 1 0 ...
$ V56      : num 0 0 0 0 0 0 0 0 0 ...
$ V57      : num 0 0 1 0 0 1 1 0 0 1 ...
$ V58      : num 0 0 0 0 0 0 0 0 0 ...
$ V55      : num 35 21 2 3 0 12 103 61 7 0 ...

```

[list output truncated]

```

- attr(*"internal.selfref")=<externalptr>
>
> train <- blogData_train; test <- fbtest
> head(train); head(test) # A
tibble: 6 x 281
  plikes checkin talking category d5 d6 d7 d8 d9 d10 d11
d12 d13 d14 d15 d16 d17 d18 d19 d20 d21 d22 d23
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 40.3 0 377 2 34.6 48.5 0 378 12 1.48 46.2 -356
2 40.3 0 377 2 34.6 48.5 0 378 12 1.48 46.2 -356
3 40.3 0 377 2 34.6 48.5 0 378 12 1.48 46.2 -356
4 40.3 0 377 2 34.6 48.5 0 378 12 1.48 46.2 -356
5 40.3 0 377 2 34.6 48.5 0 378 12 1.48 46.2 -356
6 40.3 0 377 2 34.6 48.5 0 378 12 1.48 46.2 -356
# ... with 258 more variables: d24 <dbl>, d25 <dbl>, d26 <dbl>, d27 <dbl>, d28
<dbl>, d29 <dbl>, cc1 <dbl>, cc2 <dbl>, cc3 <dbl>, cc4 <dbl>,
# cc5 <dbl>, basetime <dbl>, postlength <dbl>, postshre <dbl>, postpromo
<dbl>, Hhrs <dbl>, sun <dbl>, mon <dbl>, tue <dbl>, wed <dbl>, thu <dbl>, # fri
<dbl>, sat <dbl>, basesun <dbl>, basemon <dbl>, basetue <dbl>, basewed <dbl>,
basethu <dbl>, basefri <dbl>, basesat <dbl>, target <dbl>,
# NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA
<dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>,
# NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA
<dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>,
# NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA
<dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>,
# NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA

```

<dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, NA <dbl>, ...

| plikes checkin talking | | | | | category d5 d6 | | d7 d8 d9 | d10 d11 d12 |
|-------------------------|---------------------|----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-------------|
| d13 d14 d15 d16 d17 | d18 d19 d20 d21 | d22 d23 d24 1: | | | | | | |
| 10. 630660 | 0 17. 882992 | 1 259 5.0 0 0 4.0182760 | | | | | | |
| 10. 39679 0 0 0 235 | 0 0 1.0 3. 8172395 | 0 0 0 10. 297346 0 0 | | | | | | |
| 2: 43. 435825 | 0 75. 590485 | 0 634 20.0 0 0 15. 9985895 | | | | | | |
| 44. 56087 0 0 0 473 | 0 1 2.0 15. 4696760 | 0 0 0 44. 685085 0 0 | | | | | | |
| 3: 1. 733333 | 0 3. 043390 | 0 9 0.0 0 0 0. 7333333 | | | | | | |
| 1. 52607 0 0 0 5 | 0 1 0.0 0. 6666667 | 0 1 1. 534782 0 0 | | | | | | |
| d25 d26 d27 d28 d29 | cc1 | cc2 cc3 cc4 cc5 | basetime | postlength | | | | |
| postshre postpromo | Hhrs sun mon | tue wed thu fri sat | | | | | | |
| 1: 0 235 1 0 0 | 9. 776869 | 16. 073494 0 0 1 192 | 0 | | | | | |
| 0 5.0 0. 20103656 | 0 0 13. 948867 -229 | 0 0 0 217 | 0 | | | | | |
| 2: 0 473 1 0 0 | 40. 971790 | 70. 307840 0 0 0 479 | 0 | | | | | |
| 0 18.0 0. 52891400 | 0 0 62. 134968 -461 | 0 0 0 473 | 0 | | | | | |
| 3: 0 5 0 0 0 | 1. 133333 | 1. 820867 1 1 0 5 | 0 | | | | | |
| 0 0.0 0. 06666667 | 0 0 1. 730767 -5 | 0 0 0 4 | 0 | | | | | |
| basesun basemon basetue | basewed | basethu basefri basesat | target | V55 V56 V57 | | | | |
| V58 | V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | | | | |
| 1: 0.0 | 0 0. 2517731 | 0. 9038038 | 0 | 0 | 0 | 14 | 0 | |
| 0 0. 09438080 | 0 0. 5067316 | 0 0 0 12 | 0 | 0 | 0 | 0 | 0 | |
| 2: 0.0 | 0 0. 1932299 | 0. 4576994 | 0 | 0 | 0 | 2 | 0 | |
| 0 0. 07334273 | 0 0. 2864750 | 0 0 0 2 | 0 | 0 | 0 | 0 | 0 | |
| 3: 0.0 | 0 0. 3333333 | 0. 6992059 | 0 | 0 | 0 | 2 | 1 | |
| 0 0. 13333334 | 0 0. 3399347 | 0 0 0 1 | 0 | 0 | 0 | 0 | 0 | |
| V58 | V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| 0. 5042160 | 0 0 0 12 | 0 0 0 0. 23349700 | 0 | 0 | 0 | 0 | 0 | |
| 0. 8547111 | 0 0 0 0 | 0 0. 002454992 | 0. 6747285 | 0 | 0 | -10 | 12 | |
| 2: 0. 06770099 | 0. 2778884 | 0 0 0 0 | 0. 17630465 | 0 | 0 | 0 | 0 | |
| 0. 4297832 | 0 0 0 2 | 0 0. 005641749 | 0. 4044489 | 0 | 0 | -2 | 2 | |
| 3: 0. 13333334 | 0. 3399347 | 0 0 0 1 | 0. 20000000 | 0 | 0 | 0 | 0 | |
| 0. 4000000 | 0 0 0 1 | 0 0. 00000000 | 0. 3651484 | 0 | 0 | -1 | 1 | |
| V56 V57 V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| V57 V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| 0 0 35 0 | 0 0 35 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 2: 0 0 0 21 | 0 0 0 2 | 0 0 21 -2 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 696 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 3: 0 1 0 2 | 0 0 2 0 | 0 0 2 2 | 0 0 0 2 | 0 0 0 2 | 0 0 0 2 | 0 0 0 2 | 0 0 0 2 | |
| 0 2 0 0 | 0 2 13 1 | 0 0 8361 | 0 0 0 1 | 0 0 0 1 | 0 0 0 1 | 0 0 0 1 | 0 0 0 1 | |
| V56 V57 V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| V57 V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| 1: 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 2: 0 0 1 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 3: 0 0 1 1 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| V57 V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| 1: 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 2: 1 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 3: 1 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| V57 V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| 1: 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 2: 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 3: 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 1 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| V57 V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | V58 V55 V56 V57 | |
| 1: 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 2: 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 3: 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |
| 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | 0 0 0 0 | |

3: 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

```

V57 V58 V55 V56 V57 V58 1: 0
      4    0    0    0    0
2:    0    0    0    0    0    0
3:    0    1    0    0    0    1
      [ reached getOption("max.print") -- omitted 3 rows ]
>
> # making the data tidy by constructing single column 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 column 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))))))
>
> # a. Create a linear regression model to predict the number of comments in the next 24
hours
> # (relative to basetime)
> #install.packages(MASS)
> library(MASS)
>
> final_model <- lm(target ~ checkin + talking + d5 + d6 + d7 + d8 + d9 + d10
+ d11 +
+               d12 + d13 + d16 + d17 + d19 + d20 + d21 + d22 + d23 + d24 +
+               cc1 + cc2 + cc3 + cc4 + basetime + postshre + Hhrs + wed +
+               thu + fri + basemon + basewed, data = train)
> summary(final_model)

```

Call:

```

lm(formula = target ~ checkin + talking + d5 + d6 + d7 + d8 + d9 + d10 +
    d11 + d12 + d13 + d16 + d17 + d19 + d20 + d21 +
    d22 + d23 + d24 + cc1 + cc2 + cc3 + cc4 + basetime + postshre + Hhrs + wed +
    thu + fri + basemon + basewed, data = train)

```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|------|---------|
| -561.78 | -13.04 | -1.83 | 0.00 | 1743.64 |

Coefficients: (2 not defined because of singularities)

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|------------|------------|---------|--------------|
| (Intercept) | -4.947e-04 | 5.171e-01 | -0.001 | 0.999 |
| checkin | 3.892e-05 | 1.692e-01 | 0.000 | 1.000 |
| talking | 1.700e-04 | 1.203e-01 | 0.001 | 0.999 |
| d5 | 1.263e-05 | 1.282e-01 | 0.000 | 1.000 |
| d6 | -9.984e+02 | 6.649e+05 | -0.002 | 0.999 |
| d7 | -1.411e-03 | 5.473e-01 | -0.003 | 0.998 |
| d8 | 4.528e-04 | 3.698e+00 | 0.000 | 1.000 |
| d9 | 3.487e-05 | 2.347e-02 | 0.001 | 0.999 |
| d10 | -3.316e-04 | 1.752e-01 | -0.002 | 0.998 |
| d11 | 9.984e+02 | 6.649e+05 | 0.002 | 0.999 |
| d12 | 3.521e-04 | 3.883e-01 | 0.001 | 0.999 |
| d13 | NA | NA | NA | NA |
| d16 | 9.999e-01 | 1.997e-01 | 5.007 | 5.55e-07 *** |
| d17 | 5.831e-05 | 1.330e-01 | 0.000 | 1.000 |
| d19 | -1.190e-05 | 1.008e-02 | -0.001 | 0.999 |
| d20 | -8.603e-05 | 1.488e-01 | -0.001 | 1.000 |
| d21 | 9.984e+02 | 6.649e+05 | 0.002 | 0.999 |
| d22 | 5.252e-04 | 2.758e-01 | 0.002 | 0.998 |
| d23 | 1.633e-05 | 1.088e-02 | 0.002 | 0.999 |
| d24 | -1.133e-06 | 1.780e-02 | 0.000 | 1.000 |
| cc1 | -7.536e-03 | 1.932e+00 | -0.004 | 0.997 |
| cc2 | 1.402e-02 | 8.699e+00 | 0.002 | 0.999 |
| cc3 | 2.395e-04 | 1.436e+01 | 0.000 | 1.000 |
| cc4 | NA | NA | NA | NA |

| | | | | |
|----------|------------|-----------|--------|-------|
| basetime | -8.246e-03 | 1.027e+01 | -0.001 | 0.999 |
| postshre | 2.803e-03 | 1.443e+01 | 0.000 | 1.000 |
| Hhrs | -8.483e-04 | 8.746e-01 | -0.001 | 0.999 |
| wed | 8.755e-04 | 4.810e+00 | 0.000 | 1.000 |
| thu | 3.968e-04 | 3.294e-01 | 0.001 | 0.999 |
| fri | 4.796e-04 | 1.784e+00 | 0.000 | 1.000 |
| basemon | -2.404e-04 | 8.184e-01 | 0.000 | 1.000 |
| basewed | 4.229e-03 | 2.081e+01 | 0.000 | 1.000 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 67.27 on 52366 degrees of freedom Multiple
R-squared: 0.4976, Adjusted R-squared: 0.4973 F-statistic:
1788 on 29 and 52366 DF, p-value: < 2.2e-16

Conclusion/Interpretation:

A linear regression model is created to predict the number of comments in the next 24 hours (relative to base time) and following observation is obtained:
Residual standard error: 67.27 on 52366 degrees of freedom
Multiple R-squared: 0.4976, Adjusted R-squared: 0.4973 F-
statistic: 1788 on 29 and 52366 DF, p-value: < 2.2e-16

b. Fine tune the model and represent important features Visualize the dataset and make inferences from that.

The R-script for the given problem is as follows:

```
final_model <- lm(target ~ talking + d5 + d7 + d8 + d10 + d11 +
                  d12 + d13 + d16 + d17 + d19 + d20 + d22 + d23 +
                  cc1 + cc2 + cc3 + cc4 + basetime + postshre + Hhrs, data = train)
summary(final_model)

prediction <- predict(final_model, test)
predicted <- data.frame(cbind(actuals = test$target, prediction = prediction))
predicted$prediction <- ifelse(prediction < 0, 0, round(prediction, 0))
cor(predicted)
View(predicted)
```

The output of the R-Script (from Console window) is given as follows:

```
> final_model <- lm(target ~ talking + d5 + d7 + d8 + d10 + d11 +
+                   d12 + d13 + d16 + d17 + d19 + d20 + d22 + d23 +
+                   cc1 + cc2 + cc3 + cc4 + basetime + postshre + Hhrs,
data = train)
> summary(final_model)
```

Call:

```
lm(formula = target ~ talking + d5 + d7 + d8 + d10 + d11 + d12 + d13 + d16 +
    d17 + d19 + d20 + d22 + d23 + cc1 + cc2 + cc3 + cc4 + basetime +
    postshre + Hhrs, data = train)
```

Residuals:

| | | | | |
|-----|----|--------|----|-----|
| Min | 1Q | Median | 3Q | Max |
|-----|----|--------|----|-----|

-561.80 -13.04 -1.82 0.00 1743.64

Coefficients: (2 not defined because of singularities)

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|------------|------------|---------|-------------|
| (Intercept) | -2.666e-04 | 5.008e-01 | -0.001 | 1.000 |
| talking | -5.647e-05 | 8.142e-02 | -0.001 | 0.999 |
| d5 | 3.962e-05 | 1.220e-01 | 0.000 | 1.000 |
| d7 | -2.452e-04 | 2.637e-01 | -0.001 | 0.999 |
| d8 | 1.983e-03 | 3.565e+00 | 0.001 | 1.000 |
| d10 | 2.184e-05 | 1.151e-01 | 0.000 | 1.000 |
| d11 | 1.320e-04 | 4.381e-01 | 0.000 | 1.000 |
| d12 | 5.752e-05 | 3.379e-01 | 0.000 | 1.000 |
| d13 | NA | NA | NA | NA |
| d16 | 1.000e+00 | 1.706e-01 | 5.862 | 4.6e-09 *** |
| d17 | -8.778e-05 | 8.403e-02 | -0.001 | 0.999 |
| d19 | 1.831e-06 | 3.633e-03 | 0.001 | 1.000 |
| d20 | -1.967e-04 | 1.334e-01 | -0.001 | 0.999 |
| d22 | 1.001e-04 | 1.820e-01 | 0.001 | 1.000 |
| d23 | -2.766e-06 | 4.255e-03 | -0.001 | 0.999 |
| cc1 | -6.988e-03 | 1.420e+00 | -0.005 | 0.996 |
| cc2 | 1.500e-02 | 7.565e+00 | 0.002 | 0.998 |
| cc3 | -1.322e-03 | 1.424e+01 | 0.000 | 1.000 |
| cc4 | NA | NA | NA | NA |
| basetime | -9.812e-04 | 8.624e+00 | 0.000 | 1.000 |
| postshre | 3.307e-03 | 1.424e+01 | 0.000 | 1.000 |
| Hhrs | -1.781e-04 | 1.616e-01 | -0.001 | 0.999 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 67.26 on 52376 degrees of freedom Multiple R-squared: 0.4976, Adjusted R-squared: 0.4974 F-statistic: 2730 on 19 and 52376 DF, p-value: < 2.2e-16

```
>
> prediction <- predict(final_model, test)
> predicted <- data.frame(cbind(actuals = test$target, prediction = prediction))
> predicted$prediction <- ifelse(prediction<0, 0, round(prediction,0))
> cor(predicted)
      actuals    prediction
actuals 1.00000000 -0.03790971
prediction -0.03790971 1.00000000
> View(predicted)
```

Conclusion/Interpretation:

Residual standard error: 67.26 on 52376 degrees of freedom Multiple R-squared: 0.4976, Adjusted R-squared: 0.4974 F-statistic: 2730 on 19 and 52376 DF, p-value: < 2.2e-16

c. Interpret the summary of the linear model.

The R-script for the given problem is as follows:

```
summary(final_model)
```

The output of the R-Script (from Console window) is given as follows:

```
> summary(final_model)
```

Call:

```
lm(formula = target ~ talking + d5 + d7 + d8 + d10 + d11 + d12 + d13 + d16 +  
    d17 + d19 + d20 + d22 + d23 + cc1 + cc2 + cc3 + cc4 + basetime +  
    postshre + Hhrs, data = train)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|------|---------|
| -561.80 | -13.04 | -1.82 | 0.00 | 1743.64 |

Coefficients: (2 not defined because of singularities)

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|------------|------------|---------|-------------|
| (Intercept) | -2.666e-04 | 5.008e-01 | -0.001 | 1.000 |
| talking | -5.647e-05 | 8.142e-02 | -0.001 | 0.999 |
| d5 | 3.962e-05 | 1.220e-01 | 0.000 | 1.000 |
| d7 | -2.452e-04 | 2.637e-01 | -0.001 | 0.999 |
| d8 | 1.983e-03 | 3.565e+00 | 0.001 | 1.000 |
| d10 | 2.184e-05 | 1.151e-01 | 0.000 | 1.000 |
| d11 | 1.320e-04 | 4.381e-01 | 0.000 | 1.000 |
| d12 | 5.752e-05 | 3.379e-01 | 0.000 | 1.000 |
| d13 | NA | NA | NA | NA |
| d16 | 1.000e+00 | 1.706e-01 | 5.862 | 4.6e-09 *** |
| d17 | -8.778e-05 | 8.403e-02 | -0.001 | 0.999 |
| d19 | 1.831e-06 | 3.633e-03 | 0.001 | 1.000 |
| d20 | -1.967e-04 | 1.334e-01 | -0.001 | 0.999 |
| d22 | 1.001e-04 | 1.820e-01 | 0.001 | 1.000 |
| d23 | -2.766e-06 | 4.255e-03 | -0.001 | 0.999 |
| cc1 | -6.988e-03 | 1.420e+00 | -0.005 | 0.996 |
| cc2 | 1.500e-02 | 7.565e+00 | 0.002 | 0.998 |
| cc3 | -1.322e-03 | 1.424e+01 | 0.000 | 1.000 |
| cc4 | NA | NA | NA | NA |
| basetime | -9.812e-04 | 8.624e+00 | 0.000 | 1.000 |
| postshre | 3.307e-03 | 1.424e+01 | 0.000 | 1.000 |
| Hhrs | -1.781e-04 | 1.616e-01 | -0.001 | 0.999 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 67.26 on 52376 degrees of freedom Multiple R-squared: 0.4976, Adjusted R-squared: 0.4974 F-statistic: 2730 on 19 and 52376 DF, p-value: < 2.2e-16

Conclusion/Interpretation:

Residual standard error: 67.26 on 52376 degrees of freedom Multiple R-squared: 0.4976, Adjusted R-squared: 0.4974 F-statistic: 2730 on 19 and 52376 DF, p-value: < 2.2e-16

d. Report the test accuracy vs. the training accuracy

The R-script for the given problem is as follows:

```
# test accuracy  
round(accuracy(predicted$prediction,predicted$actuals),3)
```



```
prediction <- predict(final_model, test)
predicted <- data.frame(cbind(actuals = test$target, prediction = prediction))
```

```

predicted$prediction <- ifelse(prediction<0, 0, round(prediction,0))

min_max_accuracy <- mean(apply(predicted, 1, min) / apply(predicted, 1, max))

# training accuracy
round(accuracy(predicted$prediction,predicted$actuals),3)

prediction <- predict(final_model, train)
predicted <- data.frame(cbind(actuals = train$target, prediction = prediction))
predicted$prediction <- ifelse(prediction<0, 0, round(prediction, 0))
min_max_accuracy <- mean(apply(predicted, 1, min) / apply(predicted, 1, max))

```

The output of the R-Script (from Console window) is given as follows:

```

> # test accuracy
> round(accuracy(predicted$prediction,predicted$actuals),3) ME RMSE
              MAE      MPE MAPE
Test set -0.007 67.251 27.405 -Inf      Inf
>
> prediction <- predict(final_model, test)

> predicted <- data.frame(cbind(actuals = test$target, prediction = prediction))
> predicted$prediction <- ifelse(prediction<0, 0, round(prediction,0))
>
> min_max_accuracy <- mean(apply(predicted, 1, min) / apply(predicted, 1, max))
> # training accuracy
> round(accuracy(predicted$prediction,predicted$actuals),3) ME RMSE
              MAE      MPE MAPE
Test set -0.026 0.207 0.035 -Inf      Inf
>
> prediction <- predict(final_model, train)
> predicted <- data.frame(cbind(actuals = train$target, prediction = prediction))
> predicted$prediction <- ifelse(prediction<0, 0, round(prediction, 0))
> min_max_accuracy <- mean(apply(predicted, 1, min) / apply(predicted, 1, max))

```

Conclusion/Interpretation:

FOR TEST DATASET:

| | ME | RMSE | MAE | MPE | MAPE |
|----------|--------|--------|--------|------|------|
| Test set | -0.007 | 67.251 | 27.405 | -Inf | Inf |

FOR TRAIN DATASET

| | ME | RMSE | MAE | MPE | MAPE |
|----------|--------|-------|-------|------|------|
| Test set | -0.026 | 0.207 | 0.035 | -Inf | Inf |