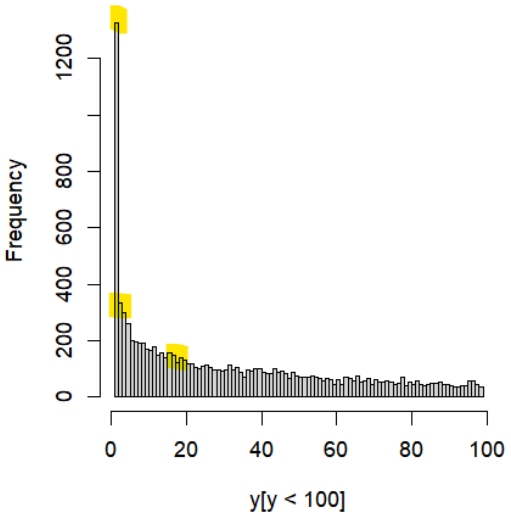
Gaia Vaccarezza 3073561

All the code is in R except for some special cases. When I switch from R to Python I will specify it.

1. Data

I started by creating the variables: month, day, hour, week_day and day_year (day of the year from 1 to 365) by extracting information from the column "opening_datetime". I did this using python, through which I also eliminated the rows that did not have this information (5 rows). The decision was due to the fact that their y are low and the other variables have no unusual labels. While for the test data I put 15/06/2021 12.00 which is the middle of the year.

After I started visualizing each variable, plotting it through histograms. For the y variable, I immediately noticed that its distribution is highly skewed with long tails (high kurtosis). It is skewed because y takes only positive values and it has a negative exponential shape. The great majority of the data concentrate from 1 to 300 (80%), some more from 400 to 1300 (reaching 98,7%) and very few observations from 1300 to 10000. These are followed by less than 20 values from 10000 to 30000 and then just 3 values: 50047, 53840 and 100800, which are much higher than all the other observations (2 times and 3 times the 4th highest value). I decided to not eliminate these extremely high values for the moment because I did not have enough information to understand whether they are outliers, measurement errors or if there was a specific trend behind them. Nevertheless, I tried fitting all the models on different samples, with and without these strange observations, to understand which is better for prediction using cross-validation error and test error. On the right, you can see the frequency of the value of y from 1 to 100 and the peak of 1.

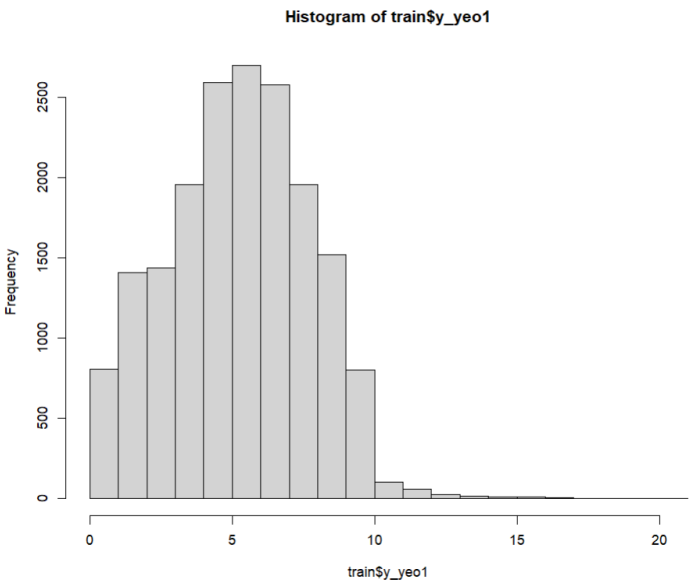


In other variables, I noticed that some labels or some values (for the continuous ones) are more frequent than others. In some cases, this characteristic is pronounced and I took it into account.

Then I plotted each independent variable against y to see if there was a particular trend or if there were some areas that were more crowded or some strange values. Firstly, I plotted all the data and then only observations with y<1300 because it was not easy to read the first graphs due to the high values of y that scaled the image. However, there are no particular trends except for products with lower code (probably older products and more problematic or

more bought) and month and day_year that show a peek at the end and the start of the year. It was useful also to understand the existence of some labels that span all the values of y and so have no effect on it.
I proceeded by plotting the boxplot of each variable in order to see if there was a label that was particularly important to predict y and I found some interesting information that I then used when deciding which dummies to add to the model. In this step, I only considered labels with more than a certain frequency (sometimes 59, sometimes 89) to not be biased by trends that may be just random or not robust.

2. Transformation of y
As I noticed the strange distribution of y I decided to apply log and square root transformation to the data. However, the histograms of the two transformations were not particularly convincing since they are still a little bit skewed and so I looked for something else. I read some articles about data that follow a negative exponential distribution and it is advised to apply the Yeo-Johnson transformation. From what I understood it is a generalization of BoxCox transformation that is a middle ground between log and sqrt transformation and one can decide the "weight" of the two by changing the parameter lambda. There is also a way to choose the best lambda by maximizing the likelihood. In my case, I decided to find the lambda by plotting the graph of the likelihood but I also try to perform some cross-validation between different levels of lambda around the value found through the graph. At the left, there is y transformed at yeo-johnson with lambda=0.9. The maximization gave a value of 0.06 but 0.09 performed slightly better with cross-validation. It has a much more "bell shape" than the log.



3. Regression
As y is a continuous variable it is a regression problem. I started building a regression in a naive manner. Firstly, I decided to work with all data, then I decided to work just with observations with y<10000 and y<1300. In both cases, for categorical variables I started by fitting y ~ as.factor(variable_name) and then I took only the 1% significant labels. However, there are some labels that are more frequent or that span the majority of the y values (for more frequent y) and this biased the results of the t-tests, making significant variables that are not. Therefore, I also used information coming from boxplots to add labels that seem to have the majority of their quantiles at a specific level of y. Of course, taking care of the length of the whiskers and outliers.

So after I used the as.factor and the boxplot I encoded the majority of the categorical variables as dummies for each label, except for cases like priority. After this, I built a regression for each variable (meaning using important or significant dummies coming from the same variable) and then I made combinations of different variables. Each time I combined two or more variables I took notes of which variables become insignificant to find out possible trends. I used plots and residual plots when it was possible to look for non-linear trends, heteroskedasticity and outliers. Then I calculated the VIF to look for the correlation between variables.
As I said before I did the same step both for different train sets and for different transformations and I used both cross-validation error and test error (uploading the result on the platform) to understand which was the best solution.
However, the results of the complete regressions, meaning the one with all significant variables were not satisfactory. Therefore, I tried to apply some best subset selections to them. It was very computationally consuming as I knew before. I chose the best dimension firstly by using R_adjusted and then calculating CV error and I used the one-standard-error rule.
Also, in this case, the cv error and the test error were high so I performed shrinkage through lasso and ridge choosing the perfect lambda (shrinkage parameter) through cross-validation.
Still, the error was high so I assumed that there was a problem with how I was choosing variables to include or more probably the model is more complex than a linear one and so I had to think of something else.

4. Tree, Random Forest and Boosting

As the majority of the variables are dummies coded as 0-1 I found that dealing with linearity through polynomial bases would not be worth it. Moreover, the complexity of the model and the large span of the values of y made me understand that there was the necessity of using trees. I wanted to behave in the same naive manner as before, building trees for each categorical variable using as.factor since the algorithm is able to understand this and use them also in more levels and not stop at the first split. However, R function for trees but also for random forests and boosting requires the number of factors to be lower than 30 and 52. Therefore I decided to proceed in a more "dirty" way. I basically divided the observations based on values of y split into 10 intervals. Each interval has approximately the same amount of data to have equal distribution but of course, this means that we have large intervals in terms of y. An exception is the last two that have fewer observations but I was done on purpose. After this split, I looked for each of the categorical variables, the list of the most frequent labels on each interval and the relative frequency. Visualizing through a table I was able to see if some labels were equally present in all intervals or more characteristic to a

```
train11 <- train[train$y>9999,]
train10 <- train[train$y>2999 & train$y<10000,]
train9 <- train[train$y>1299 & train$y<3000,]
train8 <- train[train$y>399 & train$y<1300,]
train7 <- train[train$y>184 & train$y<400,]
train6 <- train[train$y>84 & train$y<185,]
train5 <- train[train$y>39 & train$y<85,]
train4 <- train[train$y>14 & train$y<40,]
train3 <- train[train$y>3 & train$y<15,]
train2 <- train[train$y<4,]
```

specific one following the way the tree algorithm works. It was very useful even if it was quite long but it was the only way to solve the problem of the limit of factors in the R function for trees and random forests. Using the labels found and the other continuous variable I firstly build several trees and then a bigger one. However, the cross-validation error was high, and the test error was even worse probably due to the high variance of such a big tree. Therefore, I went directly to a random forest, without considering bagging as random forests do the same averaging predictions but also decorrelating trees that in this case (lots of variables probably really correlated) is very important. Indeed it is worth noting that I am fitting a model putting both some dummies of operator, work_group and service_center even if they are one a subset of the other and so they are correlated.

Random Forest gave a strong improvement and also provided important information on the importance (through measures of importance) of the single variables. Through these measures, I tried to drop some variables but unfortunately, the result was not very satisfactory so I decided to keep them. The parameters chosen were the number of variables to consider at each split equal to p/3 (=29) as it is a regression and the n of trees = 500. The first parameter comes from the theory while the second is from cross-validation doing 3 models one with tree= 100, 300, 500. To choose the last was also taken into account the % Var explained by the model. It should be noted that the % variance explained for 500 was around 40% which is quite low and this indicates that something wrong was going on. However, I will deepen this in the next paragraph.

Finally, I tried boosting, which gave the best results. Boosting learns sequentially and in this model with around 100 parameters, a high span of y and some strange outcome observation is very important. Indeed, each time it fits a tree it does it on a re-weighted version of the dataset giving more weights to observations that were wrongly predicted before. Here I tried models with different transformations, with different depths (from 1 to 4), with different n of the tree (from 200 to 5000) and with varying levels of shrinkage (0.001 and 0,01). It was using both cross-validations with 5 folds and test error from the platform I found that the best model is with depth=3, shrinkage=0.001 and ntrees=5000. This is a very slow-learner model.
Also in this case I used important measures to cut some useless variables but the model did not improve particularly both in terms of cross-validation error and test error.

Some variables that seem to be important in more than a model are: products 41, 956(for a high value of y), work group: 131, 121, 24, 22, 33 and 7, priority, channel 2, operator: 911, 2331, 1213, 20, 1223, service center 13, 8, day of the year (1 to 365), sla_contract, and hour. Among these ones, the best are:

day_year, sla_contract, channel_2. For what concerned consumers there was no clear information.

5. Some final try

As the model is not particularly predictive and for different folds gives a very different cross-validation error from 600000 to 1400000, I tried to re-think the model a bit. I went back to the tables that I have done to choose the dummies and I found that I could try to add a dummy for open and closing time instead of using hour as a continuous variable. However, even if the dummy is quite important the test error does not improve.
Then I tried to add the square of day_year as I had found that there was a higher y at first and start of the year, something that may be seen as a convex parabola.
Finally, I tried again to see the correlation against variables because I felt I was missing something and indeed the workgroup, customers, product and service center are all correlated between them and with many other variables. Therefore I tried again random forest as it decorrelates trees that may be highly correlated either due to "strong" variables or due to correlated variables. This time I exaggerated with a number of trees choosing the value of 1000 and I tried both with p/3 (30) and sqrt(p) (10) variables at each split. However, they did not lower the test error.

6. For the future

In my opinion, it should be some clustering and a re-targeting of the labels taking into account their average value of y. Moreover, one idea could be to do some oversampling of the high values of y.

7. Most important lines of code

```
install.packages("car")
library(car)
install.packages("forecast")
library(forecast)
train$y_yeo1 <- yjPower(train$y, 0.09)
hist(train$y_yeo1)
install.packages("gbm")
library("gbm")
#boosting with yeo 0.9 5000 trees and depth 3
boost = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
```

```
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract +hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)
pred = predict(boost, newdata=test)
install.packages("VGAM")
library("VGAM")
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_11.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)
```

End of the report. After you find all the code.

Code

```
#python
train.loc[1,"opening_datetime_int"] = datetime.strptime(train.loc[1,"opening_datetime"], format_data)
train.head
from datetime import datetime
#format_data = "%Y-%m-%d %H:%M:%S"
train['opening_datetime'] = pd.to_datetime(train['opening_datetime'])
train['month'] = train['opening_datetime'].dt.month
train['day'] = train['opening_datetime'].dt.day
train['hour'] = train['opening_datetime'].dt.hour
train['weekday'] = train['opening_datetime'].dt.weekday
train['day_year'] = train['opening_datetime'].dt.dayofyear

train.to_csv("C:/Users/39346/OneDrive/Desktop/train__complete_time.csv", index=True)

test['opening_datetime'] = pd.to_datetime(test['opening_datetime'])
test['month'] = test['opening_datetime'].dt.month
test['day'] = test['opening_datetime'].dt.day
test['hour'] = test['opening_datetime'].dt.hour
test['weekday'] = test['opening_datetime'].dt.weekday
test['day_year'] = test['opening_datetime'].dt.dayofyear
test.loc[(test.month.isna()), 'month'] = 6
test.loc[(test.day_year.isna()), 'day_year'] = 182
test.loc[(test.day.isna()), 'day'] = 15
test.loc[(test.weekday.isna()), 'weekday'] = 3
test.loc[(test.hour.isna()), 'hour'] = 12
test.to_csv("C:/Users/39346/OneDrive/Desktop/test (1).csv", index=True)


#r
summary(train)
attach(train)
boxplot(y)

#histogram
hist(y[y<3])
hist(y[y>2 & y<10])
hist(y[y>9 & y<30])
hist(y[y>29 & y<100])
hist(y[y>99 & y<200])
hist(y[y>199 & y<400])
hist(y[y>399 & y<1000])
hist(y[y>999 & y<3000])
hist(y[y>3000 ])
hist(y[y<100], breaks= 100)
hist(y[y>1 & y<300], breaks= 300)
hist(y[y<300], breaks= 300)
hist(y[y<1500], breaks= 20)
hist(y[1500<y & y<10000], breaks=5)
hist(y[y>10000], breaks= 5)
```

```
boxplot(product)
hist(product)
hist(product[product<350], breaks=350)

hist(product[y<1300], breaks=350)

par(mfrow=c(1,1))
hist(product[y<1300], breaks=350)
hist(product[y>1300], breaks=350)
hist(product[product>350], breaks=700)
hist(product[product>350 & product<800], breaks=450)
boxplot(work_group)
hist(work_group)
hist(work_group[work_group<70], breaks=80)
hist(work_group[work_group>70], breaks=70)
boxplot(customer)
hist(customer)
hist(customer, breaks=14000)
hist(customer[customer<10], breaks= 10)
hist(customer[customer >1 & customer<100], breaks= 99)
hist(customer[customer >100 & customer<1500], breaks= 99)
hist(customer[1500<customer & customer<10000], breaks=5)
boxplot(priority)
hist(priority[priority>1])
hist(priority)
boxplot(type)
hist(type, breaks=21)
hist(type[type>1], breaks=21)
hist(channel)
hist(operator, breaks=6140)
hist(service_center)
hist(service_mode)
hist(sla_contract)
hist(forwarded)
drop(y1)

#plot
plot(product, y)
train1 <- train[train$y<1300,]
drop(train1$w)
plot(train1$product, train1$y)
plot(train1$work_group, train1$y)
plot(train1$customer, train1$y)
plot(train1$priority, train1$y)
plot(train1$type, train1$y)
plot(train1$channel, train1$y)
plot(train1$operator, train1$y)
plot(train$channel, train$y)
plot(train1$service_center, train1$y)
plot(train1$service_mode, train1$y)
plot(train1$sla_contract, train1$y)
plot(train1$forwarded, train1$y)
```

```
plot(train_time$month, train_time$y)
plot(train$month, train$y)
plot(train_time$day, train_time$y)
par(mfrow=c(1,1))
plot(train$day_year, train$y)

àboxplot
install.packages("lattice")
library("lattice")
par(mfrow=c(2,1))
bwplot(log(train_time$y) ~ as.factor(train_time$weekday), train_time)
bwplot(train_time$y ~ as.factor(train_time$weekday), train_time)
bwplot(train_time$y ~ as.factor(train_time$month), train_time)

bwplot(log(train_time$y) ~ as.factor(train_time$month), train_time)
bwplot(train_time$y ~ as.factor(train_time$day), train_time)
bwplot(train_time$y ~ as.factor(train_time$hour), train_time)


install.packages("ggplot2")
library("ggplot2")
ggplot(train_time, aes(x = as.factor(weekday), y = log(y)))+
  geom_boxplot()+
  geom_line(stat = "hline", yintercept = "mean")


train2 <- train[train$y<5000,]
drop(train2$w)
plot(train2$product, train2$y)
plot(train2$work_group, train2$y)
plot(train2$customer, train2$y)
plot(train2$priority, train2$y)
plot(train2$type, train2$y)
plot(train2$channel, train2$y)
plot(train2$operator, train1$y)
plot(train2$service_center, train2$y)
plot(train2$service_mode, train2$y)
plot(train2$sla_contract, train2$y)
plot(train2$forwarded, train2$y)

#tables for trees
train11 <- train[train$y>9999,]
train10 <- train[train$y>2999 & train$y<10000,]
train9 <- train[train$y>1299 & train$y<3000,]
train8 <- train[train$y>399 & train$y<1300,]
train7 <- train[train$y>184 & train$y<400,]
train6 <- train[train$y>84 & train$y<185,]
train5 <- train[train$y>39 & train$y<85,]
train4 <- train[train$y>14 & train$y<40,]
train3 <- train[train$y>3 & train$y<15,]
train2 <- train[train$y<4,]
lm.fit <- lm(y~as.factor(product), data = train1)
```

```
summary(lm.fit)

table(train1$product)

install.packages('plyr')
library('plyr')
c_product <- count(train1, 'product')
c_product1 <- count(train_product1, 'product')
c_workgroup <- count(train1, 'work_group')
c_customer <- count(train1, 'customer')
c_type <- count(train1, 'type')
c_operator <- count(train1, 'operator')
c_service_center <- count(train1, 'service_center')

c_product2 <- count(train2, 'product')
c_workgroup2 <- count(train2, 'work_group')
c_customer2 <- count(train2, 'customer')
c_type2 <- count(train2, 'type')
c_operator2 <- count(train2, 'operator')
c_service_center2 <- count(train2, 'service_center')

c_product3 <- count(train3, 'product')
c_workgroup3 <- count(train3, 'work_group')
c_customer3 <- count(train3, 'customer')
c_type3 <- count(train3, 'type')
c_operator3 <- count(train3, 'operator')
c_service_center3 <- count(train3, 'service_center')

c_product4 <- count(train4, 'product')
c_workgroup4 <- count(train4, 'work_group')
c_customer4 <- count(train4, 'customer')
c_type4 <- count(train4, 'type')
c_operator4 <- count(train4, 'operator')
c_service_center4 <- count(train4, 'service_center')

c_product5 <- count(train5, 'product')
c_workgroup5 <- count(train5, 'work_group')
c_customer5 <- count(train5, 'customer')
c_type5 <- count(train5, 'type')
c_operator5 <- count(train5, 'operator')
c_service_center5 <- count(train5, 'service_center')

c_product6 <- count(train6, 'product')
c_workgroup6 <- count(train6, 'work_group')
c_customer6 <- count(train6, 'customer')
c_type6 <- count(train6, 'type')
c_operator6 <- count(train6, 'operator')
c_service_center6 <- count(train6, 'service_center')

c_product7 <- count(train7, 'product')
c_workgroup7 <- count(train7, 'work_group')
c_customer7 <- count(train7, 'customer')
```

```
c_type7 <- count(train7, 'type')
c_operator7 <- count(train7, 'operator')
c_service_center7 <- count(train7, 'service_center')

c_product8 <- count(train8, 'product')
c_workgroup8 <- count(train8, 'work_group')
c_customer8 <- count(train8, 'customer')
c_type8 <- count(train8, 'type')
c_operator8 <- count(train8, 'operator')
c_service_center8 <- count(train8, 'service_center')

c_product9 <- count(train9, 'product')
c_workgroup9 <- count(train9, 'work_group')
c_customer9 <- count(train9, 'customer')
c_type9 <- count(train9, 'type')
c_operator9 <- count(train9, 'operator')
c_service_center9 <- count(train9, 'service_center')

c_product10 <- count(train10, 'product')
c_workgroup10 <- count(train10, 'work_group')
c_customer10 <- count(train10, 'customer')
c_type10 <- count(train10, 'type')
c_operator10 <- count(train10, 'operator')
c_service_center10 <- count(train10, 'service_center')

c_product11 <- count(train11, 'product')
c_workgroup11 <- count(train11, 'work_group')
c_customer11 <- count(train11, 'customer')
c_type11 <- count(train11, 'type')
c_operator11 <- count(train11, 'operator')
c_service_center11 <- count(train11, 'service_center')


c_priority2 <- count(train2, 'priority')
c_priority3 <- count(train3, 'priority')
c_priority4 <- count(train4, 'priority')
c_priority5 <- count(train5, 'priority')
c_priority6 <- count(train6, 'priority')
c_priority7 <- count(train7, 'priority')
c_priority8 <- count(train8, 'priority')
c_priority9 <- count(train9, 'priority')
c_priority10 <- count(train10, 'priority')
c_priority11 <- count(train11, 'priority')

c_month2 <- count(train2, 'month')
c_month3 <- count(train3, 'month')
c_month4 <- count(train4, 'month')
c_month5 <- count(train5, 'month')
c_month6 <- count(train6, 'month')
c_month7 <- count(train7, 'month')
c_month8 <- count(train8, 'month')
c_month9 <- count(train9, 'month')
```

```
c_month10 <- count(train10, 'month')
c_month11 <- count(train11, 'month')

c_hour2 <- count(train2, 'hour')
c_hour3 <- count(train3, 'hour')
c_hour4 <- count(train4, 'hour')
c_hour5 <- count(train5, 'hour')
c_hour6 <- count(train6, 'hour')
c_hour7 <- count(train7, 'hour')
c_hour8 <- count(train8, 'hour')
c_hour9 <- count(train9, 'hour')
c_hour10 <- count(train10, 'hour')
c_hour11 <- count(train11, 'hour')

c_channel2 <- count(train2, 'channel')
c_channel3 <- count(train3, 'channel')
c_channel4 <- count(train4, 'channel')
c_channel5 <- count(train5, 'channel')
c_channel6 <- count(train6, 'channel')
c_channel7 <- count(train7, 'channel')
c_channel8 <- count(train8, 'channel')
c_channel9 <- count(train9, 'channel')
c_channel10 <- count(train10, 'channel')
c_channel11 <- count(train11, 'channel')

c_service_mode2 <- count(train2, 'service_mode')
c_service_mode3 <- count(train3, 'service_mode')
c_service_mode4 <- count(train4, 'service_mode')
c_service_mode5 <- count(train5, 'service_mode')
c_service_mode6 <- count(train6, 'service_mode')
c_service_mode7 <- count(train7, 'service_mode')
c_service_mode8 <- count(train8, 'service_mode')
c_service_mode9 <- count(train9, 'service_mode')
c_service_mode10 <- count(train10, 'service_mode')
c_service_mode11 <- count(train11, 'service_mode')

#sometimes I use different train with frequency of each label (using only more frequent labels and
dropping rows with non frequent labels) It is not wise but I tried
train_product <- read.csv("C:/Users/39346/OneDrive/Desktop/train_product.csv")
train_product1 <- train_product[train_product$frequency_product>89,]
bwplot(train_product1$y ~ as.factor(train_product1$product), train_product1)
lm.fit <- lm(y~as.factor(product), data = train_product1)
summary(lm.fit)
install.packages("fastDummies")
library(fastDummies)
train_product <- dummy_cols(train_product, select_columns = "product")
lm.fit <- lm(y~ product_41+ product_104 + product_158+ product_203 + product_203 + product_234+
product_276 + product_312 + product_313 + product_347 + product_457+ product_472 +
product_606, data = train_product)
summary(lm.fit)
lm.fit <- lm(log(y)~as.factor(product), data = train_product1)
summary(lm.fit)
```

```
lm.fit <- lm(log(y)~ product_50 + product_104 + product_158+ product_161 + product_203 +
product_203 + product_234+ product_276 +product_298 +  product_312 + product_313 +
product_347 +product_396+  product_457+ product_472 + product_606 + product_679, data =
train_product)
summary(lm.fit) #nuovo con log

table(train1$work_group)
train_workgroup <- read.csv("C:/Users/39346/OneDrive/Desktop/train_workgroup.csv")
train_workgroup1 <- train_workgroup[train_workgroup$frequency_workgroup>89,]
bwplot(train_workgroup1$y ~ as.factor(train_workgroup1$work_group), train_workgroup1)
lm.fit <- lm(y~as.factor(work_group), data = train_workgroup1)
summary(lm.fit)
library(fastDummies)
train_workgroup <- dummy_cols(train_workgroup, select_columns = "work_group")
lm.fit <- lm(y~ work_group_8 + work_group_9 + work_group_17 + work_group_18 + work_group_22 +
work_group_23+ work_group_24 + work_group_25 + work_group_26 + work_group_27 +
work_group_28 + work_group_29 + work_group_31 + work_group_33 + work_group_47 +
work_group_50 + work_group_53 + work_group_121 + work_group_131, data = train_workgroup)
summary(lm.fit)
lm.fit <- lm(log(y)~as.factor(work_group), data = train_workgroup1)
summary(lm.fit)
lm.fit <- lm(log(y)~ work_group_1 + work_group_17 + work_group_18  +  work_group_22 +
work_group_24 + work_group_25 + work_group_26 + work_group_27 + work_group_28 +
work_group_29 + work_group_31 + work_group_33 + work_group_47 + work_group_50 +
work_group_53 +  work_group_121 + work_group_131, data = train_workgroup)
summary(lm.fit) #nuovo con log

table(train1$customer)
train_customer <- read.csv("C:/Users/39346/OneDrive/Desktop/train_customer.csv")
train_customer1 <- train_customer[train_customer$frequency_customer>100,]
c_costumer1 <- count(train_customer1, 'customer')
bwplot(train_customer1$y ~ as.factor(train_customer1$customer), train_customer1)

lm.fit <- lm(y~as.factor(customer), data = train_customer1)
summary(lm.fit)
library(fastDummies)
train_customer <- dummy_cols(train_customer, select_columns = "customer")
lm.fit <- lm(y~ customer_90 + customer_559  + customer_1228  + customer_1314 + customer_1613 +
customer_2743 + customer_2852 + customer_3032+ customer_3752 + customer_3758 +
customer_3927 + customer_5622 + customer_5709 + customer_6967 + customer_8912 +
customer_10353 + customer_10467 + customer_10853 + customer_10893 + customer_11214, data=
train_customer)
summary(lm.fit)
lm.fit <- lm(log(y)~as.factor(customer), data = train_customer1)
summary(lm.fit)
lm.fit <- lm(log(y)~ customer_1 + customer_90 + customer_559   + customer_1228 + customer_1613
+ customer_2743 + customer_2852  + customer_3752 + customer_3758 + customer_3927 +
customer_5622 + customer_5709 + customer_10467 + customer_10853 + customer_10893 +
customer_11214, data= train_customer)
summary(lm.fit)

table(train1$priority)
```

```
lm.fit <- lm(y~as.factor(priority), data = train1)
summary(lm.fit)
bwplot(train1$y ~ as.factor(train1$priority), train1)
lm.fit <- lm(log(y)~as.factor(priority), data = train1)
summary(lm.fit)
train1 <- dummy_cols(train1, select_columns = "priority")
lm.fit <- lm(log(y)~priority_2 + priority_3, data = train1)
summary(lm.fit)

table(train1$type)
library(fastDummies)
train_type  <- read.csv("C:/Users/39346/OneDrive/Desktop/train_allfreq.csv")
train_type1 <- train_type[train_type$frequency_type>59,]
train1 <- dummy_cols(train1, select_columns = "type")
lm.fit <- lm(y~ as.factor(type), data =train1)
bwplot(train_type1$y ~ as.factor(train_type1$type) , train_type1)
lm.fit <- lm(y~ type_3 + type_5 + type_9 + type_20, data =train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ as.factor(type), data =train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ type_3 + type_5 + type_9 + type_17, data =train1)
summary(lm.fit)

table(train1$channel)
library(fastDummies)
train1 <- dummy_cols(train1, select_columns = "channel")
lm.fit <- lm(y~ as.factor(channel), data =train1)
bwplot(train1$y ~ as.factor(train1$channel) , train1)
lm.fit <- lm(y~ channel_2 + channel_5 +  channel_4, data =train1)
summary(lm.fit)
lm.fit <- lm(log(y) ~ as.factor(channel), data =train1)
lm.fit <- lm(log(y) ~ channel_2 + channel_5 + channel_4, data =train1)


table(train1$operator)
train_operator <- read.csv("C:/Users/39346/OneDrive/Desktop/train_operator.csv")
train_operator1 <- train_operator[train_operator$frequency_operator>120,]
bwplot(train_operator1$y ~ as.factor(train_operator1$operator) , train_operator1)
lm.fit <- lm(y~as.factor(operator) , data = train_operator1)
summary(lm.fit)
library(fastDummies)
train_operator <- dummy_cols(train_operator, select_columns = "operator")
lm.fit <- lm(y~ operator_155+  operator_157 +  operator_164 +  operator_166 +  operator_169 +
operator_200 +  operator_219 +  operator_219 +  operator_596 +  operator_635 +  operator_638 +
operator_899 +  operator_910 +  operator_911 +  operator_1213+  operator_1218 +  operator_1219 +
operator_1220 +  operator_1223+  operator_1564 +  operator_1565 +  operator_1941 +
operator_2331 +  operator_2736 +  operator_2738 +  operator_2742 +  operator_2744 +
operator_3651 +  operator_3652 +  operator_4174 +  operator_4198 +  operator_4202 +
operator_4206 +  operator_4207 +  operator_4220 +  operator_4802 +  operator_4823 +
operator_5455 +  operator_5456, data= train_operator)
summary(lm.fit)
lm.fit <- lm(log(y)~as.factor(operator) , data = train_operator1)
```

```
lm.fit <- lm(log(y)~ operator_155+  operator_157 +  operator_164 +  operator_166 +  operator_169 +
operator_184 + operator_198+  operator_200 +  operator_219 + operator_225 + operator_596 +
operator_635 +  operator_638 +  operator_899 +  operator_910 +  operator_911 +  operator_1213+
operator_1218 +  operator_1219 +  operator_1220 +  operator_1223+ operator_1224+
operator_1564 +  operator_1565 + operator_1567+  operator_1941 +  operator_2314 +
operator_2331 +  operator_2736 +  operator_2738 +  operator_2742 +  operator_2744 +
operator_3185 + operator_3650 + operator_3651 +  operator_3652 + operator_3653 +operator_4148
+ operator_4158 + operator_4159+ operator_4171+   operator_4174 +  operator_4198 +
operator_4201+  operator_4202 + operator_4203+  operator_4206 +  operator_4207 + operator_4210
+ operator_4211 + operator_4216 +   operator_4220 +  operator_4802 +  operator_4814+
operator_4823 +  operator_5455 +  operator_5456, data= train_operator)
summary(lm.fit) #new with log

table(train1$service_center)
library(fastDummies)
train1 <- dummy_cols(train1, select_columns = "service_center")
bwplot(train1$y ~ as.factor(train1$service_center) , train1)
lm.fit <-lm( y~ service_center_1 + service_center_3 + service_center_6 + service_center_7 +
service_center_8 + service_center_9 + service_center_10 + service_center_11 + service_center_13 +
service_center_14 + service_center_15 + service_center_17 + service_center_18 +
service_center_19, data = train1)
summary(lm.fit)
lm.fit <-lm(log(y)~ service_center_1 + service_center_3 + service_center_6 + service_center_7 +
service_center_8 + service_center_9 + service_center_10 + service_center_11 + service_center_13 +
service_center_14 + service_center_15 + service_center_17 + service_center_18 +
service_center_19, data = train1)
summary(lm.fit)
lm.fit <-lm(log(y)~ service_center_1 + service_center_3 + service_center_7 + service_center_8 +
service_center_10 + service_center_11 + service_center_13 + service_center_14, data = train1)
summary(lm.fit) #new with log

table(train1$service_mode)
lm.fit <- lm(y~as.factor(service_mode) , data = train1)
bwplot(train1$y ~ as.factor(train1$service_mode) , train1)
summary(lm.fit)
lm.fit <- lm(log(y)~as.factor(service_mode) , data = train1)
summary(lm.fit)
train1 <- dummy_cols(train, select_columns = "service_mode")
lm.fit <- lm(log(y)~ service_mode_2, data = train1)
summary(lm.fit)
#new with log


table(train1$sla_contract)
train1 <- dummy_cols(train1, select_columns = "sla_contract")
lm.fit <- lm(y~ sla_contract_1, data = train1)
lm.fit <- lm(y~as.factor(sla_contract) , data = train1)
summary(lm.fit)
lm.fit <- lm(log(y)~as.factor(sla_contract) , data = train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ sla_contract_1, data = train1)
```

```
table(train1$forwarded)
lm.fit <- lm(y~ as.factor(forwarded), data = train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ as.factor(forwarded), data = train1)
summary(lm.fit)

install.packages("fastDummies")
library(fastDummies)
train1 <- dummy_cols(train1, select_columns = "operator")
train1 <- dummy_cols(train1, select_columns = "customer")
train1 <- dummy_cols(train1, select_columns = "product")
train1 <- dummy_cols(train1, select_columns = "work_group")
train1 <- dummy_cols(train1, select_columns = "priority")
train1 <- dummy_cols(train1, select_columns = "channel")
train1 <- dummy_cols(train1, select_columns = "type")
train1 <- dummy_cols(train1, select_columns = "service_mode")
train1 <- dummy_cols(train1, select_columns = "service_center")
train1 <- dummy_cols(train1, select_columns = "sla_contract")

lm.fit <- lm(y~ channel_2 + channel_5 +  channel_4+ type_3 + type_5 + type_9 + sla_contract_1 +
as.factor(service_mode)+  as.factor(priority) + customer_90 + customer_559  + customer_1228  +
customer_1314 + customer_1613 + customer_2743 + customer_2852 + customer_3032 +
customer_3752 + customer_3752 + customer_3758 + customer_3927 + customer_5622 +
customer_5709 + customer_6967 + customer_8912 + customer_10353 + customer_10467 +
customer_10853 + customer_10893 + customer_11214+ product_41+ product_104 + product_158+
product_203 + product_234+ product_276 + product_312 + product_313 + product_347 +
product_457+ product_472 + product_606 + as.factor(forwarded) + work_group_8 + work_group_9 +
work_group_17 + work_group_18 + work_group_22 + work_group_23+ work_group_24 +
work_group_25 + work_group_26 + work_group_27 + work_group_28 + work_group_29 +
work_group_31 + work_group_33 + work_group_47 + work_group_50 + work_group_53 +
work_group_121 + work_group_131 + service_center_1 + service_center_3 + service_center_8 +
service_center_10 +  service_center_13 + service_center_14+ operator_155+  operator_157 +
operator_164 +  operator_166 +  operator_169 +  operator_200 +  operator_219 +  operator_219 +
operator_596 +  operator_635 +  operator_638 +  operator_899 +  operator_910 +  operator_911 +
operator_1213+  operator_1218 +  operator_1219 +  operator_1220 +  operator_1223+
operator_1564 +  operator_1565 +  operator_1941 +  operator_2331 +  operator_2736 +
operator_2738 +  operator_2742 +  operator_2744 +  operator_3651 +  operator_3652 +
operator_4174 +  operator_4198 +  operator_4202 +  operator_4206 +  operator_4207 +
operator_4220 +  operator_4802 +  operator_4823 +  operator_5455 +  operator_5456, data= train1)
summary(lm.fit)
summary(lm.fit)

lm.fit <- lm(y ~ channel_2 + channel_5 +  channel_4+ type_3 + type_5 + sla_contract_1 + priority_3 +
customer_90 + customer_559 + customer_2743 + customer_11214+ product_41+ product_104 +
product_234 + product_312 + product_347 + product_457 + as.factor(forwarded) + work_group_8 +
work_group_9 + work_group_22 + work_group_25 + work_group_26 + work_group_27 +
work_group_28 + work_group_53 + work_group_121 + work_group_131 + service_center_3 +
service_center_14+  operator_164 +  operator_169  +  operator_219 +  operator_219 +  operator_596
+  operator_635 +  operator_638 +  operator_899 +  operator_910 +  operator_911 +  operator_1213+
operator_1218 +  operator_1220 +  operator_1223+  operator_1564 +  operator_1565 +
operator_1941 +  operator_2331 +  operator_2738 +  operator_2742 +  operator_2744 +
```

```
operator_4202 +  operator_4206 +  operator_4220 +  operator_4802 +  operator_4823 +
operator_5455 +  operator_5456, data= train1)
summary(lm.fit)

lm.fit <- lm(y ~ channel_2 + channel_5 +  channel_4+ type_3 + type_5 + sla_contract_1 + priority_3 +
customer_90 + customer_559 + customer_2743 + customer_11214+ product_41+ product_104 +
product_234 + product_312 + product_347 + product_457 + as.factor(forwarded) + work_group_8 +
work_group_9 + work_group_22 + work_group_25 + work_group_26 + work_group_27 +
work_group_28 + work_group_53 + work_group_121 + work_group_131 + service_center_3 +
service_center_14+  operator_164 +  operator_169  +  operator_219 +  operator_219 +  operator_596
+  operator_635 +  operator_638 +  operator_899 +  operator_910 +  operator_911 +  operator_1213+
operator_1218 +  operator_1220 +  operator_1223+  operator_1564 +  operator_1565 +
operator_1941 +  operator_2331 +  operator_2738 +  operator_2742 +  operator_2744 +
operator_4202 +  operator_4206 +  operator_4220 +  operator_4802 +  operator_4823 +
operator_5455 +  operator_5456, data= train1)
summary(lm.fit)

lm.fit <- lm(y ~ channel_2 + channel_5 +  channel_4+ type_3 + type_5 + sla_contract_1 + priority_3 +
customer_90 + customer_559 + customer_2743 + customer_11214+ product_41+ product_104 +
product_234 + product_312 + product_347 + product_457 + work_group_8 + work_group_9 +
work_group_22 + work_group_25 + work_group_26 + work_group_27 + work_group_28 +
work_group_53 + work_group_121 + work_group_131 + service_center_3 + service_center_14+
operator_164 +  operator_169  +  operator_219 +  operator_219 +  operator_596 +  operator_635 +
operator_638 +  operator_899 +  operator_910 +  operator_911 +  operator_1213+  operator_1218 +
operator_1220 +  operator_1223+  operator_1564 +  operator_1565 +  operator_1941 +
operator_2331 +  operator_2738 +  operator_2744 +  operator_4202 +  operator_4206 +
operator_4220 +  operator_4802 +  operator_4823 +  operator_5455 +  operator_5456, data= train1)
summary(lm.fit)

lm.fit <- lm(y ~ channel_2 + channel_5 +  channel_4+ type_3 + type_5 + sla_contract_1 + priority_3 +
customer_90 + customer_559 + customer_2743 + product_41+ product_104 + product_234 +
product_347 + product_457 + work_group_8 + work_group_9 + work_group_22 + work_group_25 +
work_group_26 + work_group_28 + work_group_53 + work_group_121 + work_group_131 +
service_center_3 + service_center_14+  operator_164 +  operator_169  +  operator_219 +
operator_219 +  operator_596 +  operator_635 +  operator_638 +  operator_899 +  operator_910 +
operator_911 +  operator_1213+  operator_1218  +  operator_1223+  operator_1564 +
operator_1565 +  operator_1941 +  operator_2331 +  operator_2744 +  operator_4202 +
operator_4206 +  operator_4220 +  operator_4802 +  operator_4823 +  operator_5455 +
operator_5456, data= train1)
summary(lm.fit)

lm.fit <- lm(log(y) ~ channel_2 + channel_5 +  channel_4+ type_3 + type_5 + sla_contract_1 +
priority_3 + customer_90 + customer_559 + customer_2743 + product_41+ product_104 +
product_234 + product_347 + product_457 + work_group_8 + work_group_9 + work_group_22 +
work_group_25 + work_group_26 + work_group_28 + work_group_53 + work_group_121 +
work_group_131 + service_center_3 + service_center_14+  operator_164 +  operator_169  +
operator_219 +  operator_219 +  operator_596 +  operator_635 +  operator_638 +  operator_899 +
operator_910 +  operator_911 +  operator_1213+  operator_1218  +  operator_1223+  operator_1564
+  operator_1565 +  operator_1941 +  operator_2331 +  operator_2744 +  operator_4202 +
operator_4206 +  operator_4220 +  operator_4802 +  operator_4823 +  operator_5455 +
operator_5456, data= train1)
summary(lm.fit)
```

```
test <- read.csv("C:/Users/39346/OneDrive/Desktop/test (1).csv")
test1 <- read.csv("C:/Users/39346/OneDrive/Desktop/test_time.csv")
test <- dummy_cols(test, select_columns = "operator")
test <- dummy_cols(test, select_columns = "customer")
test <- dummy_cols(test, select_columns = "product")
test <- dummy_cols(test, select_columns = "work_group")
test<- dummy_cols(test, select_columns = "priority")
test<- dummy_cols(test, select_columns = "type")
test<- dummy_cols(test, select_columns = "channel")
test<- dummy_cols(test, select_columns = "service_center")
test <- dummy_cols(test, select_columns = "service_mode")
test <- dummy_cols(test, select_columns = "sla_contract")
test <- dummy_cols(test, select_columns = "hour")
drop(test$hour)
test$hour_18[test$hour.isna] <- 0
test$day_year[test$day_year.isna] <- 182

train_time <- dummy_cols(train_time, select_columns = "weekday")
train_time <- dummy_cols(train_time, select_columns = "hour")
train_time <- dummy_cols(train_time, select_columns = "weekday")
lm.pred <- predict(lm.fit, test)
ciao <- (exp(lm.pred))
lm.pred[lm.pred <1] = 1
lm.pred

ciao <- ((lm.pred)^2)
ciao
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/reg4_log.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)
write.table(lm.pred, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/reg3_y.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)


lm.fit <- lm(log(y) ~ channel_2 + channel_5 +  channel_4+ type_3 + sla_contract_1 + priority_3 +
customer_90 + customer_559 + customer_2743 + product_41+ product_104 + product_234 +
product_347 + product_457 + work_group_22 + work_group_25  + work_group_28 + work_group_53
+ work_group_121 + work_group_131 + service_center_3 + service_center_14+  operator_164 +
operator_169  +  operator_219 +  operator_219 +  operator_596 +  operator_635 +  operator_638 +
operator_910 +  operator_911 +  operator_1213+  operator_1218  +  operator_1223+  operator_1564
+  operator_1565 +  operator_1941 +  operator_2331 +  operator_2744 +  operator_4202 +
operator_4206 +  operator_4802 +  operator_4823 +  operator_5455 +  operator_5456, data= train1)
summary(lm.fit)

lm.fit <- lm(log(y) ~ channel_2 + channel_5 +  channel_4+ type_3 + sla_contract_1 + priority_3 +
customer_90 + customer_559 + customer_2743 + product_41+ product_104 + product_234 +
product_347 + product_457 + work_group_22 + work_group_25  + work_group_28 +
work_group_121 + work_group_131 + service_center_3 + service_center_14+  operator_164 +
operator_169  +  operator_219 +  operator_219 +  operator_596 +  operator_635 +  operator_638 +
```

```
operator_910 +  operator_911 +  operator_1213+  operator_1218  +  operator_1223+  operator_1564
+  operator_1565 +  operator_2331 +  operator_2744 +  operator_4202 +  operator_4206 +
operator_4802 +  operator_4823 +  operator_5455 +  operator_5456, data= train1)
summary(lm.fit)

lm.fit <- lm(sqrt(y) ~ channel_2 + channel_5 +  channel_4+ type_3 + type_5 + sla_contract_1 +
priority_3 + customer_90 + customer_559 + customer_2743 + product_41+ product_104 +
product_234 + product_347 + product_457 + work_group_8 + work_group_9 + work_group_22 +
work_group_25 + work_group_26 + work_group_28 + work_group_53 + work_group_121 +
work_group_131 + service_center_3 + service_center_14+  operator_164 +  operator_169  +
operator_219 +  operator_219 +  operator_596 +  operator_635 +  operator_638 +  operator_899 +
operator_910 +  operator_911 +  operator_1213+  operator_1218  +  operator_1223+  operator_1564
+  operator_1565 +  operator_1941 +  operator_2331 +  operator_2744 +  operator_4202 +
operator_4206 +  operator_4220 +  operator_4802 +  operator_4823 +  operator_5455 +
operator_5456, data= train1)
summary(lm.fit)

#new with log
lm.fit <- lm(log(y)~ product_50 + product_104 + product_158+ product_161 + product_203 +
product_203 + product_234+ product_276 +product_298 +  product_312 + product_313 +
product_347 +product_396+  product_457+ product_472 + product_606 + product_679+
work_group_1 + work_group_17 + work_group_18  +  work_group_22 + work_group_24 +
work_group_25 + work_group_26 + work_group_27 + work_group_28 + work_group_29 +
work_group_31 + work_group_33 + work_group_47 + work_group_50 + work_group_53 +
work_group_121 + work_group_131 + customer_1 + customer_90 + customer_559   +
customer_1228 + customer_1613 + customer_2743 + customer_2852  + customer_3752 +
customer_3758 + customer_3927 + customer_5622 + customer_5709 + customer_10467 +
customer_10853 + customer_10893 + customer_11214 + priority_2 + priority_3 + type_3 + type_5 +
type_9 + type_17 +channel_1 + channel_2 + channel_5 + channel_4 + operator_155+  operator_157
+  operator_164 +  operator_166 +  operator_169 +  operator_184 + operator_198+  operator_200 +
operator_219 + operator_225 + operator_596 +  operator_635 +  operator_638 +  operator_899 +
operator_910 +  operator_911 +  operator_1213+  operator_1218 +  operator_1219 +  operator_1220
+  operator_1223+ operator_1224+  operator_1564 +  operator_1565 + operator_1567+
operator_1941 +  operator_2314 + operator_2331 +  operator_2736 +  operator_2738 +
operator_2742 +  operator_2744 +  operator_3185 + operator_3650 + operator_3651 +
operator_3652 + operator_3653 +operator_4148 + operator_4158 + operator_4159+ operator_4171+
operator_4174 +  operator_4198 + operator_4201+  operator_4202 + operator_4203+  operator_4206
+  operator_4207 + operator_4210 + operator_4211 + operator_4216 +   operator_4220 +
operator_4802 +  operator_4814+   operator_4823 +  operator_5455 +  operator_5456+
service_center_1 + service_center_3 + service_center_7 + service_center_8 + service_center_10 +
service_center_11 + service_center_13 + service_center_14 + service_mode_2 + sla_contract_1 +
as.factor(forwarded), data = train1)
summary(lm.fit)  #Radj 0.41

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_53 +  work_group_121 + work_group_131 + customer_90 + customer_559   +
customer_1228 +  + customer_2743 + customer_3927 + customer_5622 + customer_5709 +
customer_10467 + customer_10853 + priority_3 + type_3 + type_9 + type_17 +channel_1 +
channel_2 + channel_5 + channel_4 + operator_155+  operator_157 +  operator_164 +  operator_169
+  operator_184 + operator_198+  operator_200 +  operator_219 + operator_225 +  operator_635 +
operator_638 +  operator_910 +  operator_911 +  operator_1213+  operator_1218 +  operator_1219 +
```

```
operator_1223+ operator_1224+  operator_1564 +  operator_1565 + operator_1567+  operator_2314
+ operator_2331 +  operator_2736  +  operator_2742  + operator_3650 + operator_3651 +
operator_3652 + operator_3653 +operator_4148 + operator_4158 + operator_4171+   operator_4174
+  operator_4198 + operator_4201+  operator_4202 +  operator_4206  + operator_4211 +
operator_4216 +  operator_4802 +  operator_4814+   operator_4823 +  operator_5455 +
operator_5456+ service_center_1 + service_center_3 + service_center_7 + service_center_8 +
service_center_10 + service_center_11 + service_center_14  + sla_contract_1 + forwarded, data =
train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_53 +  work_group_121 + work_group_131 + customer_90 + customer_559   +
customer_2743  + customer_5622+ customer_10853 + priority_3 + type_3 + type_9 + type_17
+channel_1 + channel_2 + channel_5 + channel_4 + operator_155+  operator_157 +  operator_164 +
operator_169  + operator_198+  operator_200 +  operator_219 + operator_225 +  operator_635 +
operator_638 +  operator_910 +  operator_911 +  operator_1213+  operator_1218 +  operator_1219 +
operator_1223+  operator_1564 +  operator_1565 + operator_1567+  operator_2314 + operator_2331
+  operator_2736  +  operator_2742  + operator_3650 +  operator_3652 + operator_3653
+operator_4148 + operator_4158+   operator_4174 +  operator_4198 + operator_4201+
operator_4202 +  operator_4206 + operator_4216 +  operator_4802 +   operator_4823 +
operator_5455 +  operator_5456+ service_center_1 + service_center_3 + service_center_7 +
service_center_8 + service_center_10 + service_center_14  + sla_contract_1, data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_121 + customer_90 + customer_559   + customer_2743  + customer_5622+
customer_10853 + priority_3 + type_3 + type_9 +channel_1 + channel_2 + channel_5 +
operator_155+  operator_157 +  operator_164 +  operator_169  + operator_198+  operator_200 +
operator_219 + operator_225 +  operator_635 +  operator_638 +  operator_910 +  operator_911 +
operator_1213+  operator_1218 +  operator_1219 +  operator_1223+  operator_1564 +
operator_1565 + operator_1567+  operator_2314 + operator_2331 +  operator_2736  +
operator_2742  + operator_3650 +  operator_3652 + operator_3653 +operator_4148 +
operator_4158+   operator_4174 +  operator_4198 + operator_4201+  operator_4202 +
operator_4206 + operator_4216 +  operator_4802 +   operator_4823 +  operator_5455 +
operator_5456+ service_center_1 + service_center_3 + service_center_7 + service_center_8 +
service_center_14  + sla_contract_1, data = train1)
summary(lm.fit) #Radj 0.407


plot(train$product, train$work_group)
plot(train$product, train$customer)
plot(train$product, train$priority)

#interaction with data above certain value
train_freq <- read.csv("C:/Users/39346/OneDrive/Desktop/train_allfreq.csv")
train_freq1<- train_freq[ train_freq$frequency_workgroup>89,]
lm.fit <- lm(log(y)~ as.factor(product)*as.factor(work_group), data= train_freq1)
summary(lm.fit)
train_freq1 <- dummy_cols(train_freq1, select_columns = "product")
lm.fit <- lm(log(y)~ product_234*as.factor(work_group), data = train_freq1)
```

```
lm.fit <- lm(log(y)~ product_234*type_5 + product_234*type_3 + product_234*type_1 +
product_234*type_20,  data = train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ product_313*type_5 + product_313*type_3 + product_313*type_1 +
product_313*type_20,  data = train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ product_104*type_5 + product_104*type_3 + product_104*type_1 +
product_104*type_20,  data = train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ product_298*type_5 + product_298*type_3 + product_298*type_1 +
product_298*type_20,  data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234*priority_3, data=train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ product_313*priority_3, data=train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ product_104*priority_3, data=train1)
summary(lm.fit)
lm.fit <- lm(log(y)~ product_298*priority_3, data=train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_53 +  work_group_121 + work_group_131 + customer_90 + customer_559   +
customer_1228 +  + customer_2743 + customer_3927 + customer_5622 + customer_5709 +
customer_10467 + customer_10853 + priority_3 + type_3 + type_9 + type_17 +channel_1 +
channel_2 + channel_5 + channel_4 + operator_155+  operator_157 +  operator_164 +  operator_169
+  operator_184 + operator_198+  operator_200 +  operator_219 + operator_225 +  operator_635 +
operator_638 +  operator_910 +  operator_911 +  operator_1213+  operator_1218 +  operator_1219 +
operator_1223+ operator_1224+  operator_1564 +  operator_1565 + operator_1567+  operator_2314
+ operator_2331 +  operator_2736  +  operator_2742  + operator_3650 + operator_3651 +
operator_3652 + operator_3653 +operator_4148 + operator_4158 + operator_4171+   operator_4174
+  operator_4198 + operator_4201+  operator_4202 +  operator_4206  + operator_4211 +
operator_4216 +  operator_4802 +  operator_4814+   operator_4823 +  operator_5455 +
operator_5456+ service_center_1 + service_center_3 + service_center_7 + service_center_8 +
service_center_10 + service_center_11 + service_center_14  + sla_contract_1 + forwarded, data =
train1)
summary(lm.fit)


#time
train_time <- read.csv("C:/Users/39346/OneDrive/Desktop/train_time.csv")
train_time <- dummy_cols(train_time, select_columns = "operator")
train_time <- dummy_cols(train_time, select_columns = "customer")
train_time <- dummy_cols(train_time, select_columns = "product")
train_time <- dummy_cols(train_time, select_columns = "work_group")
train_time <- dummy_cols(train_time, select_columns = "priority")
train_time <- dummy_cols(train_time, select_columns = "channel")
train_time <- dummy_cols(train_time, select_columns = "type")
train_time <- dummy_cols(train_time, select_columns = "service_mode")
```

```
train_time <- dummy_cols(train_time, select_columns = "service_center")
train_time <- dummy_cols(train_time, select_columns = "sla_contract")
train_time <- dummy_cols(train_time, select_columns = "hour")
train_time <- dummy_cols(train_time, select_columns = "weekday")

lm.fit <- lm(log(y)~ product_50 + product_104 + product_158+ product_161 + product_203 +
product_203 + product_234+ product_276 +product_298 +  product_312 + product_313 +
product_347 +product_396+  product_457+ product_472 + product_606 + product_679+
work_group_1 + work_group_17 + work_group_18  +  work_group_22 + work_group_24 +
work_group_25 + work_group_26 + work_group_27 + work_group_28 + work_group_29 +
work_group_31 + work_group_33 + work_group_47 + work_group_50 + work_group_53 +
work_group_121 + work_group_131 + customer_1 + customer_90 + customer_559   +
customer_1228 + customer_1613 + customer_2743 + customer_2852  + customer_3752 +
customer_3758 + customer_3927 + customer_5622 + customer_5709 + customer_10467 +
customer_10853 + customer_10893 + customer_11214 + priority_2 + priority_3 + type_3 + type_5 +
type_9 + type_17 +channel_1 + channel_2 + channel_5 + channel_4 + operator_155+  operator_157
+  operator_164 +  operator_166 +  operator_169 +  operator_184 + operator_198+  operator_200 +
operator_219 + operator_225 + operator_596 +  operator_635 +  operator_638 +  operator_899 +
operator_910 +  operator_911 +  operator_1213+  operator_1218 +  operator_1219 +  operator_1220
+  operator_1223+ operator_1224+  operator_1564 +  operator_1565 + operator_1567+
operator_1941 +  operator_2314 + operator_2331 +  operator_2736 +  operator_2738 +
operator_2742 +  operator_2744 +  operator_3185 + operator_3650 + operator_3651 +
operator_3652 + operator_3653 +operator_4148 + operator_4158 + operator_4159+ operator_4171+
operator_4174 +  operator_4198 + operator_4201+  operator_4202 + operator_4203+  operator_4206
+  operator_4207 + operator_4210 + operator_4211 + operator_4216 +   operator_4220 +
operator_4802 +  operator_4814+   operator_4823 +  operator_5455 +  operator_5456+
service_center_1 + service_center_3 + service_center_7 + service_center_8 + service_center_10 +
service_center_11 + service_center_13 + service_center_14 + service_mode_2 + sla_contract_1 +
as.factor(forwarded) + month +day + hour + weekday , data = train_time)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_121 + customer_90 + customer_559   + customer_2743  + customer_5622+
customer_10853 + priority_3 + type_3 + type_9 + channel_2 + channel_5 + operator_155+
operator_157 +  operator_164 +  operator_169  + operator_198+  operator_200 +  operator_219 +
operator_225 +  operator_635 +  operator_638 +  operator_910 +  operator_911 +  operator_1213+
operator_1218 +  operator_1219 +  operator_1223+  operator_1564 +  operator_1565 +
operator_1567+  operator_2314 + operator_2331 +  operator_2736  +  operator_2742  +
operator_3650 +  operator_3652 + operator_3653 +operator_4148 + operator_4158+   operator_4174
+  operator_4198 + operator_4201+  operator_4202 +  operator_4206 + operator_4216 +
operator_4802 +   operator_4823 +  operator_5455 +  operator_5456++ service_center_3 +
service_center_7 + service_center_8 + service_center_14  + sla_contract_1 + month + day + hour +
weekday, data = train_time)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_121 + customer_90 + customer_559   + customer_2743  + customer_5622+
customer_10853 + priority_3 + type_3 + type_9 + channel_2 + channel_5 + operator_155+
operator_157 +  operator_164 +  operator_169  + operator_198+  operator_200 +  operator_219 +
operator_225 +  operator_635 +  operator_638 +  operator_910 +  operator_911 +  operator_1213+
```

```
operator_1218 +  operator_1219 +  operator_1223+  operator_1564 +  operator_1565 +
operator_1567+  operator_2314 + operator_2331 +  operator_2736  +  operator_2742  +
operator_3650 +  operator_3652 + operator_3653 +operator_4148 + operator_4158+   operator_4174
+  operator_4198 + operator_4201+  operator_4202 +  operator_4206 + operator_4216 +
operator_4802 +   operator_4823 +  operator_5455 +  operator_5456++ service_center_3 +
service_center_7 + service_center_8 + service_center_14  + sla_contract_1 + month +as.factor(
hour) + weekday, data = train_time)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_121 + customer_90 + customer_559   + customer_2743  + customer_5622+
customer_10853 + priority_3 + type_3 + type_9 + channel_2 + channel_5 + operator_155+
operator_157 +  operator_164 +  operator_169  + operator_198+  operator_200 +  operator_219 +
operator_225 +  operator_635 +  operator_638 +  operator_910 +  operator_911 +  operator_1213+
operator_1218 +  operator_1219 +  operator_1223+  operator_1564 +  operator_1565 +
operator_1567+  operator_2314 + operator_2331 +  operator_2736  +  operator_2742  +
operator_3650 +  operator_3652 + operator_3653 +operator_4148 + operator_4158+   operator_4174
+  operator_4198 + operator_4201+  operator_4202 +  operator_4206 + operator_4216 +
operator_4802 +   operator_4823 +  operator_5455 +  operator_5456++ service_center_3 +
service_center_7 + service_center_8 + service_center_14  + sla_contract_1 + month + weekday_6+
hour_18 + hour_19, data = train_time)
summary(lm.fit) #0.4179

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_121 + customer_90 + customer_559   + customer_2743  + customer_5622+
customer_10853 + priority_3 + type_3 + type_9 + channel_2 + channel_5 + operator_155+
operator_157 +  operator_164 +  operator_169  + operator_198+  operator_200 +  operator_219 +
operator_225 +  operator_635 +  operator_638 +  operator_910 +  operator_911 +  operator_1213+
operator_1218 +  operator_1219 +  operator_1223+  operator_1564 +  operator_1565 +
operator_1567+  operator_2314 + operator_2331 +  operator_2736  +  operator_2742  +
operator_3650 +  operator_3652 + operator_3653 +operator_4148 + operator_4158+   operator_4174
+  operator_4198 + operator_4201+  operator_4202 +  operator_4206 + operator_4216 +
operator_4802 +   operator_4823 +  operator_5455 +  operator_5456++ service_center_3 +
service_center_7 + service_center_8 + service_center_14  + sla_contract_1 + month + weekday_6+
hour_18 + hour_19 + forwarded, data = train_time)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_121 + customer_90 + customer_559   + customer_2743  + customer_5622+
customer_10853 + priority_3 + type_3 + type_9 + channel_2 + channel_5 + operator_155+
operator_157 +  operator_164 +  operator_169  + operator_198+  operator_200 +  operator_219 +
operator_225 +  operator_635 +  operator_638 +  operator_910 +  operator_911 +  operator_1213+
operator_1218 +  operator_1219 +  operator_1223+  operator_1564 +  operator_1565 +
operator_1567+  operator_2314 + operator_2331 +  operator_2736  +  operator_2742  +
operator_3650 +  operator_3652 + operator_3653 +operator_4148 + operator_4158+   operator_4174
+  operator_4198 + operator_4201+  operator_4202 +  operator_4206 + operator_4216 +
operator_4802 +   operator_4823 +  operator_5455 +  operator_5456++ service_center_3 +
service_center_7 + service_center_8 + service_center_14  + sla_contract_1 + month + weekday_6+
hour_18 + hour_19 + forwarded, data = train_time)
```

```
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_121 + customer_90 + customer_559   + customer_2743  + customer_5622+
customer_10853 + priority_3 + type_3 + type_9 + channel_2 + channel_5 + operator_155+
operator_157 +  operator_164 +  operator_169  + operator_198+  operator_200 +  operator_219 +
operator_225 +  operator_635 +  operator_638 +  operator_910 +  operator_911 +  operator_1213+
operator_1218 +  operator_1219 +  operator_1223+  operator_1564 +  operator_1565 +
operator_1567+  operator_2314 + operator_2331 +  operator_2736  +  operator_2742  +
operator_3650 +  operator_3652 + operator_3653 +operator_4148 + operator_4158+   operator_4174
+  operator_4198 + operator_4201+  operator_4202 +  operator_4206 + operator_4216 +
operator_4802 +   operator_4823 +  operator_5455 +  operator_5456++ service_center_3 +
service_center_7 + service_center_8 + service_center_14  + sla_contract_1 + month + weekday_6+
hour_18 + hour_19 + forwarded, data = train_time)
summary(lm.fit)

lm.fit <- lm(log(y)~ hour_0 + hour_1 + hour_5 + hour_6 + hour_7 + hour_13+ hour_18 + hour_19 +
hour_22 + hour_23, data= train_time)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_234+ product_312 +product_347 + product_457+ product_472 +
work_group_1 + work_group_17 + work_group_18  + work_group_28  + work_group_31 +
work_group_121 + customer_90 + customer_559   + customer_2743  + customer_5622+
customer_10853 + priority_3 + type_3 + type_9 + channel_2 + channel_5 + operator_155+
operator_157 +  operator_164 +  operator_169  + operator_198+  operator_200 +  operator_219 +
operator_225 +  operator_635 +  operator_638 +  operator_910 +  operator_911 +  operator_1213+
operator_1218 +  operator_1219 +  operator_1223+  operator_1564 +  operator_1565 +
operator_1567+  operator_2314 + operator_2331 +  operator_2736  +  operator_2742  +
operator_3650 +  operator_3652 + operator_3653 +operator_4148 + operator_4158+   operator_4174
+  operator_4198 + operator_4201+  operator_4202 +  operator_4206 + operator_4216 +
operator_4802 +   operator_4823 +  operator_5455 +  operator_5456++ service_center_3 +
service_center_7 + service_center_8 + service_center_14  + sla_contract_1 + month + weekday_6+
hour_18 + hour_19 + hour_13 + forwarded, data = train_time)
summary(lm.fit)

lm.fit <- lm(log(y)~ po , data = train1)
summary(lm.fit)

#new data from boxplot

lm.fit <- lm(log(y)~ product_104 + product_203+ product_234+ product_312 +product_347 +
product_457, data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~  work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +
work_group_47 +  work_group_50  +  work_group_121, data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ customer_90 + customer_559  +customer_1228 + customer_1613  +
customer_2743  + customer_3752 + customer_3758 + customer_5709 +  customer_5622+
customer_10467 + customer_10853 , data= train1)
```

```
summary(lm.fit)

lm.fit <- lm(log(y)~ type_1, data = train1)
summary(lm.fit)

priority_3 , data = train1)

channel_2+ channel_4 + channel_5 , data = train1)

#union of these
lm.fit <- lm(log(y)~ product_104 + product_203+ product_234+ product_312 +product_347 +
product_457 +  work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +
work_group_47 +  work_group_50  +  work_group_121, data =train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_104 + product_203+ product_234+ product_312 +product_347 +
product_457 + customer_90 + customer_559  +customer_1228 + customer_1613  + customer_2743
+ customer_3752 + customer_3758 + customer_5709 +  customer_5622+ customer_10467 +
customer_10853 , data= train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_104 + product_203+ product_234+ product_312 +product_347 + type_1 ,
data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_104 + product_203+ product_234+ product_312 +product_347 + priority_3
, data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_104 + product_203+ product_234+ product_312 +product_347 +
channel_2+ channel_4 + channel_5 , data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~  work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +
work_group_47 +  work_group_50  +  work_group_121 +customer_90 + customer_559
+customer_1228 + customer_1613  + customer_2743  + customer_3752 + customer_3758 +
customer_5709 +  customer_5622+ customer_10467 + customer_10853 , data= train1)
summary(lm.fit)

lm.fit <- lm(log(y)~  work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +
work_group_47 +  work_group_50  +  work_group_121 + type_1 , data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +
work_group_47 +  work_group_50  +  work_group_121  + priority_3 , data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +
work_group_47 +  work_group_50  +  work_group_121  + channel_2+ channel_4 + channel_5 , data
= train1)
summary(lm.fit)
```

```
lm.fit <- lm(log(y)~ operator_155 + operator_157  +  operator_166 +  operator_200 +  operator_219 +
operator_638 + operator_899+  operator_910 +  operator_911 +  operator_1213+  operator_1219 +
operator_1223+  operator_1564 +  operator_1565 + operator_1941+ operator_2331 + operator_3651
+   operator_3652 + operator_4158+   operator_4174 +  operator_4198 + operator_4201+
operator_4202 +  operator_4206 + operator_4216 + operator_4823 +  operator_5455 , data = train1)
summary(lm.fit) #all stat sign

lm.fit <- lm(log(y)~ operator_155 +  operator_166 +  operator_200 + operator_899+  operator_910 +
operator_911 +  operator_1213+  operator_1219 +  operator_1223 +  operator_1565 + operator_2331
+ operator_3651 +   operator_4174 +  operator_4206 , data = train1)
summary(lm.fit) #low r adj

lm.fit <- lm(log(y)~ product_104 + product_203+ product_234+ product_312 +product_347 +
product_457 +operator_155 + operator_157  +  operator_166 +  operator_200 +  operator_219 +
operator_638 + operator_899+  operator_910 +  operator_911 +  operator_1213+  operator_1219 +
operator_1223+  operator_1564 +  operator_1565 + operator_1941+ operator_2331 + operator_3651
+   operator_3652 + operator_4158+   operator_4174 +  operator_4198 + operator_4201+
operator_4202 +  operator_4206 + operator_4216 + operator_4823 +  operator_5455 , data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_104 + product_203+ product_234+ product_312 +product_347 +
product_457+ operator_155 +  operator_166 +  operator_200 + operator_899+  operator_910 +
operator_911 +  operator_1213+  operator_1219 +  operator_1223 +  operator_1565 + operator_2331
+ operator_3651 +   operator_4174 +  operator_4206 , data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~  work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +
work_group_47 +  work_group_50  +  work_group_121 + operator_155 + operator_157  +
operator_166 +  operator_200 +  operator_219 +  operator_638 + operator_899+  operator_910 +
operator_911 +  operator_1213+  operator_1219 +  operator_1223+  operator_1564 +  operator_1565
+ operator_1941+ operator_2331 + operator_3651 +   operator_3652 + operator_4158+
operator_4174 +  operator_4198 + operator_4201+  operator_4202 +  operator_4206 +
operator_4216 + operator_4823 +  operator_5455 , data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +
work_group_47 +  work_group_50  +  work_group_121 + operator_155 +  operator_166 +
operator_200 + operator_899+  operator_910 +  operator_911 +  operator_1213+  operator_1219 +
operator_1223 +  operator_1565 + operator_2331 + operator_3651 +   operator_4174 +
operator_4206 , data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ operator_155 +  operator_166 +  operator_200 + operator_899+  operator_910 +
operator_911 +  operator_1213+  operator_1219 +  operator_1223 +  operator_1565 + operator_2331
+ operator_3651 +   operator_4174 +  operator_4206 + customer_90 + customer_559
+customer_1228 + customer_1613  + customer_2743  + customer_3752 + customer_3758 +
customer_5709 +  customer_5622+ customer_10467 + customer_10853 , data= train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ operator_155 + operator_157  +  operator_166 +  operator_200 +  operator_219 +
operator_638 + operator_899+  operator_910 +  operator_911 +  operator_1213+  operator_1219 +
operator_1223+  operator_1564 +  operator_1565 + operator_1941+ operator_2331 + operator_3651
```

```
+   operator_3652 + operator_4158+   operator_4174 +  operator_4198 + operator_4201+
operator_4202 +  operator_4206 + operator_4216 + operator_4823 +  operator_5455 + customer_90
+ customer_559  +customer_1228 + customer_1613  + customer_2743  + customer_3752 +
customer_3758 + customer_5709 +  customer_5622+ customer_10467 + customer_10853 , data=
train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ operator_155 + operator_157  +  operator_166 +  operator_200 +  operator_219 +
operator_638 + operator_899+  operator_910 +  operator_911 +  operator_1213+  operator_1219 +
operator_1223+  operator_1564 +  operator_1565 + operator_1941+ operator_2331 + operator_3651
+   operator_3652 + operator_4158+   operator_4174 +  operator_4198 + operator_4201+
operator_4202 +  operator_4206 + operator_4216 + operator_4823 +  operator_5455 + type_1, data
= train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ operator_155 +  operator_166 +  operator_200 + operator_899+  operator_910 +
operator_911 +  operator_1213+  operator_1219 +  operator_1223 +  operator_1565 + operator_2331
+ operator_3651 +   operator_4174 +  operator_4206 +channel_2+ channel_4 + channel_5 , data =
train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ service_center_1 + service_center_3 + service_center_8 + service_center_10,
data= train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_104 + product_203+ product_234+ product_312 +product_347 +
product_457 +service_center_1 + service_center_3 + service_center_8 + service_center_10, data=
train1)
summary(lm.fit)

lm.fit <- lm(log(y)~  work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +
work_group_47 +  work_group_50  +service_center_1 + service_center_3 + service_center_8 +
service_center_10, data= train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ operator_155 +  operator_166 +  operator_200 + operator_899+  operator_910 +
operator_911 +  operator_1213+  operator_1219 +  operator_1223 +  operator_1565 + operator_2331
+ operator_3651 +   operator_4174 +  operator_4206 + service_center_1 + service_center_3 +
service_center_8 + service_center_10 , data = train1)
summary(lm.fit)

lm.fit <- lm(log(y)~ product_104 + product_203+ product_234 +product_347 + product_457 +
work_group_18 +  work_group_24 +  work_group_25 +  work_group_26 +  work_group_47 +
work_group_50  +service_center_1 + service_center_3 + service_center_8 + service_center_10 +
customer_90 + customer_559  +customer_1228 + customer_1613  + customer_2743  +
customer_3752 + customer_3758 + customer_5709 +  customer_5622+ customer_10467 +
customer_10853 +priority_3+ type_1+ weekday_6 + hour_13 + hour_18 + hour_19 + channel_2+
channel_4 + channel_5 + operator_155+ operator_166 +  operator_200 + operator_899+
operator_910 +  operator_911 +  operator_1213+  operator_1219 +  operator_1223 +  operator_1565
+ operator_2331 + operator_3651 +   operator_4174 +  operator_4206 + service_center_1 +
service_center_3 + service_center_8 + service_center_10+ service_mode_2+ sla_contract +
forwarded , data = train_time)
```

```
summary(lm.fit)

lm.fit <- lm.fit <- lm(log(y)~product_104 + product_203+ product_234 +product_347 + product_457 +
work_group_18 +  work_group_24 + work_group_25 +  work_group_26 +  work_group_50
+service_center_1 + service_center_3 + service_center_8 + service_center_10 + customer_90 +
customer_559  +customer_1228 + customer_1613  + customer_2743  + customer_3752 +
customer_3758 + customer_5709 +  customer_5622+ customer_10467 + customer_10853
+priority_3+ type_1+ month + weekday_6 + hour_13 + hour_18 + hour_19 + channel_2+ channel_4 +
channel_5 + operator_155+ operator_166 + operator_899+  operator_910 +  operator_911 +
operator_1213+  operator_1219 +  operator_1223 +  + operator_2331 + operator_3651 +
operator_4174 +  operator_4206 + service_center_1 + service_center_3 + service_center_8 +
service_center_10+ service_mode_2+ sla_contract + forwarded , data = train_time)
summary(lm.fit)

lm.fit <- lm.fit <- lm(log(y)~product_104 + product_203+ product_234 +product_347 + product_457 +
work_group_18 +  work_group_24 + work_group_25 +  work_group_26 +  work_group_50
+service_center_1 + service_center_3 + service_center_8 + service_center_10 + customer_90 +
customer_559  +customer_1228 + customer_1613  + customer_2743  + customer_3752 +
customer_3758 + customer_5709 +  customer_5622+ customer_10467 + customer_10853
+priority_3+ type_1+ month + weekday_6 + hour_13 + hour_18 + hour_19 + channel_2+ channel_4 +
channel_5 + operator_155+ operator_166 + operator_899+  operator_910 +  operator_911 +
operator_1213+  operator_1219 +  operator_1223 +  + operator_2331 + operator_3651 +
operator_4174 +  operator_4206 + service_center_1 + service_center_3 + service_center_8 +
service_center_10+ service_mode_2+ sla_contract + forwarded , data = train_time)
summary(lm.fit) #39.17

lm.fit <-  lm(log(y)~product_104 + product_203+ product_234 +product_347 + product_457 +
work_group_18 +  work_group_24 + work_group_25 +  work_group_26 +  work_group_50
+service_center_1 + service_center_3 + service_center_8 + service_center_10 + customer_90 +
customer_559  +customer_1228 + customer_1613  + customer_2743  + customer_3752 +
customer_3758 + customer_5709 +  customer_5622+ customer_10467 + customer_10853
+priority_3+ type_1+ month + day_year:month+ weekday_6 + hour_13 + hour_18 + hour_19 +
channel_2+ channel_4 + channel_5 + operator_155+ operator_166 + operator_899+  operator_910 +
operator_911 +  operator_1213+  operator_1219 +  operator_1223 +  + operator_2331 +
operator_3651 +   operator_4174 +  operator_4206 + service_center_1 + service_center_3 +
service_center_8 + service_center_10+ service_mode_2+ sla_contract + forwarded , data =
train_time)
summary(lm.fit) #39.31 interaction

lm.fit <- lm.fit <- lm(log(y)~product_104 + product_203+ product_234 +product_347 + product_457 +
work_group_18 +  work_group_24 + work_group_25 +  work_group_26 +  work_group_50
+service_center_1 + service_center_3 + service_center_8 + service_center_10 + customer_90 +
customer_559  +customer_1228 + customer_1613  + customer_2743  + customer_3752 +
customer_3758 + customer_5709 +  customer_5622+ customer_10467 + customer_10853
+priority_3+ type_1+ type_3+  month + day_year:month+ weekday_6 + hour_13 + hour_18 + hour_19
+ channel_2+ channel_4 + channel_5 + operator_155+ operator_166 + operator_899+  operator_910
+  operator_911 +  operator_1213+  operator_1219 +  operator_1223 +  + operator_2331 +
operator_3651 +   operator_4174 +  operator_4206 + service_center_1 + service_center_3 +
service_center_8 + service_center_10+ service_mode_2+ sla_contract + forwarded , data =
train_time)
summary(lm.fit)
```

```
#best subset selection
install.packages("leaps")
library("leaps")
lm.fit <- regsubsets(log(y)~product_104 + product_203+ product_234 +product_347 + product_457 +
work_group_18 +  work_group_24 + work_group_25 +  work_group_26 +  work_group_50
+service_center_1 + service_center_3 + service_center_8 + service_center_10 + customer_90 +
customer_559  +customer_1228 + customer_1613  + customer_2743  + customer_3752 +
customer_3758 + customer_5709 +  customer_5622+ customer_10467 + customer_10853
+priority_3+ type_1+ type_3+  month + day_year:month+ weekday_6 + hour_13 + hour_18 + hour_19
+ channel_2+ channel_4 + channel_5 + operator_155+ operator_166 + operator_899+  operator_910
+  operator_911 +  operator_1213+  operator_1219 +  operator_1223 +  + operator_2331 +
operator_3651 +   operator_4174 +  operator_4206 + service_center_1 + service_center_3 +
service_center_8 + service_center_10+ service_mode_2+ sla_contract + forwarded , data =
train_time,really.big=T, nvmax=40)
summary(lm.fit)

reg.summary = summary(lm.fit)
reg.summary$adjr2

#tree random forest and boosting
#import modified dataset with month day hour weekday day_year(1-365)
train <- read.csv("C:/Users/39346/OneDrive/Desktop/train__complete_time.csv")
train_10000 <- train[train$y<10000,]
train_1300 <- train[train$y <1300, ]

#rescaling
install.packages("car")
library(car)
install.packages("forecast")
library(forecast)

y_yeo <- yjPower(y, 0.15)
hist(y_yeo)
y_yeo <- yjPower(train$y, 0.15)
hist(y_yeo)
y_yeo <- yjPower(y, 0.09)
hist(y_yeo)
y_yeo1 <- yjPower(train$y, 0.065)
hist(y_yeo1)
y_yeo <- yjPower(train$y, 0.055)
par(mfrow=c(1,2))
train$y_yeo <- yjPower(train$y, 0.06)
hist(train$y_yeo)
train$y_yeo1 <- yjPower(train$y, 0.09)
hist(train$y_yeo1)
train$y_yeo2 <- yjPower(train$y, 0.1)
hist(train$_yeo2)

y1 <- train_10000$y
hist(y1)
y_log1 <- log(y1)
hist(y_log1)
```

```
y_sqrt1 <- sqrt(y1)
hist(y_sqrt1)
y_yeo1 <- yjPower(y1, 0.1)
hist(y_yeo1)
y_yeo1 <- yjPower(y1, 0.075)
hist(y_yeo1)
y_yeo1 <- yjPower(y1, 0.065)
hist(y_yeo1)
train_10000$y_yeo1 <- yjPower(train_10000$y, 0.055)
hist(train_10000$y_yeo1) #first try
a <- boxCox(y1~1, family="yjPower", plotit = TRUE)


y2 <- train_1300$y
hist(y2)
y_log2 <- log(y2)
hist(y_log2)
y_sqrt1 <- sqrt(y1)
hist(y_sqrt1)
y_yeo2 <- yjPower(y2, 0.15)
hist(y_yeo2)

#to transform back
install.packages("VGAM")
library("VGAM")
y_yeo_try <- yeo.johnson(y1, 0.055)
hist(y_yeo_try)
y_try <- yeo.johnson(y_yeo_try, 0.055, inverse=TRUE)
hist(y_try)

#introduce dummy
install.packages("fastDummies")
library(fastDummies)
train  <- dummy_cols(train , select_columns = "operator")
train  <- dummy_cols(train , select_columns = "customer")
train  <- dummy_cols(train , select_columns = "product")
train  <- dummy_cols(train , select_columns = "work_group")
train  <- dummy_cols(train , select_columns = "priority")
train  <- dummy_cols(train , select_columns = "channel")
train  <- dummy_cols(train , select_columns = "type")
train  <- dummy_cols(train , select_columns = "service_mode")
train  <- dummy_cols(train , select_columns = "service_center")
train  <- dummy_cols(train , select_columns = "sla_contract")
train  <- dummy_cols(train , select_columns = "month")
train  <- dummy_cols(train , select_columns = "hour")

#tree with all data
install.packages("tree")
library(tree)

#product
```

```
tree = tree(y~ product_313 + product_203 + product_298 + product_347 + product_457 + product_41
+ product_12 + product_221 + product_158 + product_50 + product_27 + product_679 + product_956
+ product_168 , data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) # y 1

tree = tree(y_yeo~ product_313 + product_203 + product_298 + product_347 + product_457 +
product_41 + product_12 + product_221 + product_158 + product_50 + product_27 + product_679 +
product_956 + product_168 , data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #yeo 2 41

tree = tree(log(y)~ product_313 + product_203 + product_298 + product_347 + product_457 +
product_41 + product_12 + product_221 + product_158 + product_50 + product_27 + product_679 +
product_956 + product_168 , data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0)#log 2 41

#work_group
tree = tree(y~ work_group_131 + work_group_35 + work_group_29 + work_group_7 +
work_group_31 + work_group_53 + work_group_27 + work_group_26 + work_group_22 +
work_group_121 + work_group_24 + work_group_50 + work_group_20  + work_group_24  +
work_group_33 , data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0)  #y 1

tree = tree(y_yeo~ work_group_131 + work_group_35 + work_group_29 + work_group_7 +
work_group_31 + work_group_53 + work_group_27 + work_group_26 + work_group_22 +
work_group_121 + work_group_24 + work_group_50 + work_group_20  + work_group_24  +
work_group_33 , data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0)  #yeo 3 29-131

tree = tree(log(y)~ work_group_131 + work_group_35 + work_group_29 + work_group_7 +
work_group_31 + work_group_53 + work_group_27 + work_group_26 + work_group_22 +
work_group_121 + work_group_24 + work_group_50 + work_group_20  + work_group_24  +
work_group_33 , data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #log 2 29

#customer
tree = tree(y~ customer_5622 + customer_3752 + customer_5709 + customer_3758 +
customer_10467 + customer_3752 + customer_10853 + customer_3032 + customer_1+
customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + customer_3950, data = train)
summary(tree)
```

```
plot(tree)
text(tree, pretty = 0) #y 1 10674

tree = tree(y_yeo~ customer_5622 + customer_3752 + customer_5709 + customer_3758 +
customer_10467 + customer_3752 + customer_10853 + customer_3032 + customer_1+
customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + customer_3950, data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #6 y_yeo 5622 5709 3758 3752 10467

tree = tree(log(y)~ customer_5622 + customer_3752 + customer_5709 + customer_3758 +
customer_10467 + customer_3752 + customer_10853 + customer_3032 + customer_1+
customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + customer_3950, data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #6 log 5622 5709 3758 3752 10467

#priority
tree = tree(y~ priority, data=train)
summary(tree) #1

tree = tree(y_yeo~ priority, data=train)
summary(tree) #1

tree = tree(log(y)~ priority, data=train)
summary(tree) #1

#type
tree = tree(y~ as.factor(type), data=train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #y 2

tree = tree(y_yeo~ type_3 + type_5 + type_20 + type_9 +type_17, data=train)
summary(tree) #yeo 1

tree = tree(log(y)~ type_3 + type_5 + type_20 + type_9 +type_17, data=train)
summary(tree) #log 1

#channel
tree = tree(y~ channel_2 + channel_5, data=train)
summary(tree) #y 1

tree = tree(y_yeo~ channel_2 + channel_5, data=train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #yeo 2 2

tree = tree(log(y)~ channel_2 + channel_5, data=train)
summary(tree)
```

```
plot(tree)
text(tree, pretty = 0) #log 2 2

#operator
tree = tree(y~ operator_155 + operator_911 + operator_3652 + operator_4174 + operator_899 +
operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 + operator_910 +
operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 + operator_2742 +
operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828 + operator_20 +
operator_4804 + operator_225 , data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #y 1 node

tree = tree(y_yeo ~ operator_155 + operator_911 + operator_3652 + operator_4174 + operator_899 +
operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 + operator_910 +
operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 + operator_2742 +
operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828 + operator_20 +
operator_4804 + operator_225 , data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #yeo 2 node (155)

tree = tree(log(y)~ operator_155 + operator_911 + operator_3652 + operator_4174 + operator_899 +
operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 + operator_910 +
operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 + operator_2742 +
operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828 + operator_20 +
operator_4804 + operator_225 , data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #log 3 node (155, 911)

#service_center
tree = tree(y~ service_center_11 + service_center_13 + service_center_8 + service_center_19 +
service_center_1 + service_center_3 + service_center_7 + service_center_15, data = train )
summary(tree) #y 1

tree = tree(y_yeo~ service_center_11 + service_center_13 + service_center_8 + service_center_19 +
service_center_1 + service_center_3 + service_center_7 + service_center_15, data = train )
summary(tree)
plot(tree)
text(tree, pretty = 0) #yeo 2 13

tree = tree(log(y)~ service_center_11 + service_center_13 + service_center_8 + service_center_19 +
service_center_1 + service_center_3 + service_center_7 + service_center_15, data = train )
summary(tree)
plot(tree)
text(tree, pretty = 0) #log 2 13

#service mode
tree = tree(y~ service_mode_2 + service_mode_3, data = train)
summary(tree) #1
```

```
tree = tree(y_yeo~ service_mode_2 + service_mode_3, data = train)
summary(tree) #1

tree = tree(log(y)~ service_mode_2 + service_mode_3, data = train)
summary(tree) #1

#sla_contract

#forwarded

#month
tree = tree(y~ month_1 + month_2 + month_4 + month_12 + month_6, data = train)
summary(tree)

tree = tree(y_yeo~ month_1 + month_2 + month_4 + month_12 + month_6, data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #yeo 2 1

tree = tree(log(y)~ month_1 + month_2 + month_4 + month_12 + month_6, data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #log 2 1

#hour
tree = tree(y ~ hour_13 + hour_18 + hour_19 + hour_11 , data = train)
summary(tree)

tree = tree(y_yeo ~ hour_13 + hour_18 + hour_19 + hour_11 , data = train)
summary(tree)

tree = tree(log(y) ~ hour_13 + hour_18 + hour_19 + hour_11 , data = train)
summary(tree)

#total tree
tree = tree(y ~ product_313 + product_203 + product_298 + product_347 + product_457 + product_41
+ product_12 + product_221 + product_158 + product_50 + product_27 + product_679 + product_956
+ product_168 + work_group_131 + work_group_35 + work_group_29 + work_group_7 +
work_group_31 + work_group_53 + work_group_27 + work_group_26 + work_group_22 +
work_group_121 + work_group_24 + work_group_50 + work_group_20  + work_group_24  +
work_group_33 + customer_5622 + customer_3752 + customer_5709 + customer_3758 +
customer_10467 + customer_3752 + customer_10853 + customer_3032 + customer_1+
customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + customer_3950 + priority + type_3 + type_5 + type_20 + type_9
+type_17 + channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174
+ operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220
+ operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour_13 + hour_18 + hour_19 + hour_11 + month_1
+ month_2 + month_4 + month_12 + month_6 + service_mode_2 + service_mode_3, data = train)
```

```
summary(tree)
plot(tree)
text(tree, pretty = 0) #y 2 customer_10674

tree = tree(y_yeo ~ product_313 + product_203 + product_298 + product_347 + product_457 +
product_41 + product_12 + product_221 + product_158 + product_50 + product_27 + product_679 +
product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 + work_group_7
+ work_group_31 + work_group_53 + work_group_27 + work_group_26 + work_group_22 +
work_group_121 + work_group_24 + work_group_50 + work_group_20  + work_group_24  +
work_group_33 + customer_5622 + customer_3752 + customer_5709 + customer_3758 +
customer_10467 + customer_3752 + customer_10853 + customer_3032 + customer_1+
customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + customer_3950 + priority + type_3 + type_5 + type_20 + type_9
+type_17 + channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174
+ operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220
+ operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour_13 + hour_18 + hour_19 + hour_11 + month_1
+ month_2 + month_4 + month_12 + month_6 + service_mode_2 + service_mode_3, data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #yeo channel_2 sla_contract operator_2331 service_center13

tree = tree(log(y) ~ product_313 + product_203 + product_298 + product_347 + product_457 +
product_41 + product_12 + product_221 + product_158 + product_50 + product_27 + product_679 +
product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 + work_group_7
+ work_group_31 + work_group_53 + work_group_27 + work_group_26 + work_group_22 +
work_group_121 + work_group_24 + work_group_50 + work_group_20  + work_group_24  +
work_group_33 + customer_5622 + customer_3752 + customer_5709 + customer_3758 +
customer_10467 + customer_3752 + customer_10853 + customer_3032 + customer_1+
customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + customer_3950 + priority + type_3 + type_5 + type_20 + type_9
+type_17 + channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174
+ operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220
+ operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour_13 + hour_18 + hour_19 + hour_11 + month_1
+ month_2 + month_4 + month_12 + month_6 + service_mode_2 + service_mode_3, data = train)
summary(tree)
plot(tree)
text(tree, pretty = 0) #log 4 channel_2 service_center8 sla_contract

#random forest
install.packages("randomForest")
library(randomForest)

rf = randomForest(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
```

```
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + customer_3950 + priority + type_3 + type_5 + type_20 + type_9
+type_17 + channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174
+ operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220
+ operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour_13 + hour_18 + hour_19 + hour_11 + month_1
+ month_2 + month_4 + month_12 + month_6 + service_mode_2 + service_mode_3, data = train,
ntree=100, importance=TRUE)
rf #day_year channel_2 service_center13 sla_contract hour11 customer_5622
importance(rf)

#boosting
install.packages("gbm")
library("gbm")

boost = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + customer_3950 + priority + type_3 + type_5 + type_20 + type_9
+type_17 + channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174
+ operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220
+ operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=500, interaction.depth=3 )
summary(boost)

boost = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + customer_3950 + priority + type_3 + type_5 + type_20 + type_9
+type_17 + channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174
```

```
+ operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220
+ operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=500, interaction.depth=3 )
summary(boost)

boost = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)

boost = gbm(log(y) ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)

#5 fold cross validation
#assuming that the dataset in a randomic order i split it in 5 folds (17995/5=3599) and so create 5
train set
test1 = train[train$X<3601,]
test2 = train[3600<train$X & train$X<7201,]
```

```
test3 = train[7200<train$X & train$X<10801,]
test4 = train[10800<train$X & train$X<14401,]
test5 = train[14400<train$X,] #see than some a missing info
train1 = train[train$X<14401,] #without 5 fold
train2 = train[train$X<10801 | 14400<train$X,] #without 4 fold
train3 = train[train$X<7201 | 10800<train$X,] #without 3
train4 = train[train$X<3601 | 7200<train$X,] #without 2
train5 = train[train$X> 3600, ] #without1

#yeo 0.06 all data all variable n.trees=1000, interaction.depth=1
#hour and month together
x <- c(1:5)
boost1 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train1, distribution =  "gaussian", n.trees=1000, interaction.depth=1 )
summary(boost1)
pred1 = predict(boost1, newdata=test5)
scale1 <- yeo.johnson(pred1, 0.06, inverse=TRUE)
mean1 = mean((scale1-test5$y)^2)
mean1

boost2 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train2, distribution =  "gaussian", n.trees=1000, interaction.depth=1 )
```

```
summary(boost2)
pred2 = predict(boost2, newdata=test4)
scale2 <- yeo.johnson(pred2, 0.06, inverse=TRUE)
mean2 = mean((scale2-test4$y)^2)
mean2

boost3 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train3, distribution =  "gaussian", n.trees=1000, interaction.depth=1 )
summary(boost3)
pred3 = predict(boost3, newdata=test3)
scale3 <- yeo.johnson(pred3, 0.06, inverse=TRUE)
mean3 = mean((scale3-test3$y)^2)
mean3

boost4 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train4, distribution =  "gaussian", n.trees=1000, interaction.depth=1 )
summary(boost4)
pred4 = predict(boost4, newdata=test2)
scale4 <- yeo.johnson(pred4, 0.06, inverse=TRUE)
mean4 = mean((scale4-test2$y)^2)
mean4
```

```
boost5 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train5, distribution =  "gaussian", n.trees=1000, interaction.depth=1 )
summary(boost5)
pred5 = predict(boost5, newdata=test1)
scale5 <- yeo.johnson(pred5, 0.06, inverse=TRUE)
mean5 = mean((scale5-test1$y)^2)
mean5

a = mean1 + mean2 + mean3 +  mean4 + mean5
a/5


#yeo 0.06 all data all variable n.trees=1000, interaction.depth=2
#hour and month together
x <- c(1:5)
boost1 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train1, distribution =  "gaussian", n.trees=1000, interaction.depth=2 )
summary(boost1)
pred1 = predict(boost1, newdata=test5)
scale1 <- yeo.johnson(pred1, 0.06, inverse=TRUE)
mean1 = mean((scale1-test5$y)^2)
mean1
```

```
boost2 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train2, distribution =  "gaussian", n.trees=1000, interaction.depth=2 )
summary(boost2)
pred2 = predict(boost2, newdata=test4)
scale2 <- yeo.johnson(pred2, 0.06, inverse=TRUE)
mean2 = mean((scale2-test4$y)^2)
mean2

boost3 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train3, distribution =  "gaussian", n.trees=1000, interaction.depth=2 )
summary(boost3)
pred3 = predict(boost3, newdata=test3)
scale3 <- yeo.johnson(pred3, 0.06, inverse=TRUE)
mean3 = mean((scale3-test3$y)^2)
mean3

boost4 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
```

```
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train4, distribution =  "gaussian", n.trees=1000, interaction.depth=2 )
summary(boost4)
pred4 = predict(boost4, newdata=test2)
scale4 <- yeo.johnson(pred4, 0.06, inverse=TRUE)
mean4 = mean((scale4-test2$y)^2)
mean4

boost5 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train5, distribution =  "gaussian", n.trees=1000, interaction.depth=2 )
summary(boost5)
pred5 = predict(boost5, newdata=test1)
scale5 <- yeo.johnson(pred5, 0.06, inverse=TRUE)
mean5 = mean((scale5-test1$y)^2)
mean5

a = mean1 + mean2 + mean3 +  mean4 + mean5
a/5

#yeo 0.06 all data all variable n.trees=1000 interaction.depth=3
#hour and month together
boost1 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
```

```
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train1, distribution =  "gaussian", n.trees=1000, interaction.depth=3 )
summary(boost1)
pred1 = predict(boost1, newdata=test5)
scale1 <- yeo.johnson(pred1, 0.06, inverse=TRUE)
mean1 = mean((scale1-test5$y)^2)
mean1

boost2 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train2, distribution =  "gaussian", n.trees=1000, interaction.depth=3 )
summary(boost2)
pred2 = predict(boost2, newdata=test4)
scale2 <- yeo.johnson(pred2, 0.06, inverse=TRUE)
mean2 = mean((scale2-test4$y)^2)
mean2

boost3 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
```

```
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train3, distribution =  "gaussian", n.trees=1000, interaction.depth=3 )
summary(boost3)
pred3 = predict(boost3, newdata=test3)
scale3 <- yeo.johnson(pred3, 0.06, inverse=TRUE)
mean3 = mean((scale3-test3$y)^2)
mean3

boost4 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train4, distribution =  "gaussian", n.trees=1000, interaction.depth=3 )
summary(boost4)
pred4 = predict(boost4, newdata=test2)
scale4 <- yeo.johnson(pred4, 0.06, inverse=TRUE)
mean4 = mean((scale4-test2$y)^2)
mean4

boost5 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train5, distribution =  "gaussian", n.trees=1000, interaction.depth=3 )
summary(boost5)
```

```
pred5 = predict(boost5, newdata=test1)
scale5 <- yeo.johnson(pred5, 0.06, inverse=TRUE)
mean5 = mean((scale5-test1$y)^2)
mean5

a = mean1 + mean2 + mean3 +  mean4 + mean5
a/5


#yeo 0.06 all data all variable n.trees=3000 interaction.depth=3
#hour and month together
boost1 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train1, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost1)
pred1 = predict(boost1, newdata=test5)
scale1 <- yeo.johnson(pred1, 0.06, inverse=TRUE)
mean1 = mean((scale1-test5$y)^2)
mean1

boost2 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train2, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost2)
```

```
pred2 = predict(boost2, newdata=test4)
scale2 <- yeo.johnson(pred2, 0.06, inverse=TRUE)
mean2 = mean((scale2-test4$y)^2)
mean2

boost3 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train3, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost3)
pred3 = predict(boost3, newdata=test3)
scale3 <- yeo.johnson(pred3, 0.06, inverse=TRUE)
mean3 = mean((scale3-test3$y)^2)
mean3

boost4 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train4, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost4)
pred4 = predict(boost4, newdata=test2)
scale4 <- yeo.johnson(pred4, 0.06, inverse=TRUE)
mean4 = mean((scale4-test2$y)^2)
mean4
```

```
boost5 = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train5, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost5)
pred5 = predict(boost5, newdata=test1)
scale5 <- yeo.johnson(pred5, 0.06, inverse=TRUE)
mean5 = mean((scale5-test1$y)^2)
mean5

a = mean1 + mean2 + mean3 +  mean4 + mean5
a/5


#yeo 0.09 all data all variable n.trees=3000 interaction.depth=3
#hour and month together
boost1 = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train1, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost1)
pred1 = predict(boost1, newdata=test5)
scale1 <- yeo.johnson(pred1, 0.06, inverse=TRUE)
mean1 = mean((scale1-test5$y)^2)
mean1
```

```
boost2 = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train2, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost2)
pred2 = predict(boost2, newdata=test4)
scale2 <- yeo.johnson(pred2, 0.06, inverse=TRUE)
mean2 = mean((scale2-test4$y)^2)
mean2

boost3 = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train3, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost3)
pred3 = predict(boost3, newdata=test3)
scale3 <- yeo.johnson(pred3, 0.06, inverse=TRUE)
mean3 = mean((scale3-test3$y)^2)
mean3

boost4 = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
```

```
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train4, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost4)
pred4 = predict(boost4, newdata=test2)
scale4 <- yeo.johnson(pred4, 0.06, inverse=TRUE)
mean4 = mean((scale4-test2$y)^2)
mean4

boost5 = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train5, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost5)
pred5 = predict(boost5, newdata=test1)
scale5 <- yeo.johnson(pred5, 0.06, inverse=TRUE)
mean5 = mean((scale5-test1$y)^2)
mean5

a = mean1 + mean2 + mean3 +  mean4 + mean5
a/5

#yeo1 0.09  all data all variable n.trees=1000 interaction.depth=4 shrinkage 0.01
boost = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
```

```
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=1000, interaction.depth=4, shrinkage= 0.01 )
summary(boost)
pred = predict(boost, newdata=test)
ciao <- yeo.johnson(pred5, 0.09, inverse=TRUE)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
hist(ciao)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_yeo09_4level_s001.txt", append = TRUE, sep = "\n", dec = ".", row.names =
FALSE, col.names= FALSE)


fh

#yeo 1 con depth 3 e 5000 3 MENO VAR -
boost = gbm(y_yeo1 ~ day_year + product_203 + product_298 + product_41 + product_956 +
product_168 + work_group_131 + work_group_35 + work_group_29 + work_group_7 +
work_group_31 + work_group_53  + work_group_121 + work_group_24 + work_group_50  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_1 + customer_10674 + customer_927  +
customer_918 + priority + type_3 + channel_2 + channel_5 + operator_155 + operator_911  +
operator_2331+ operator_1213 + operator_1223 +  operator_2744 + operator_200 + operator_2742 +
+ operator_1218 + operator_20 + operator_4804 + operator_225 + service_center_13 +
service_center_8 + service_center_19 + service_center_3 + service_center_7+ forwarded +
sla_contract + hour  + service_mode_2 + service_mode_3, data = train, distribution =  "gaussian",
n.trees=5000, interaction.depth=3 )
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
hist(ciao)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_yeo09_lessvar.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE,
col.names= FALSE)


#log all data all variable n.trees=3000 interaction.depth=3
#hour and month together
boost1 = gbm(log(y) ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
```

```
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train1, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost1)
pred1 = predict(boost1, newdata=test5)
scale1 = exp(pred1)
mean1 = mean((scale1-test5$y)^2)
mean1

boost2 = gbm(log(y) ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train2, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost2)
pred2 = predict(boost2, newdata=test4)
scale2 = exp(pred2)
mean2 = mean((scale2-test4$y)^2)
mean2

boost3 = gbm(log(y) ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
```

```
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train3, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost3)
pred3 = predict(boost3, newdata=test3)
scale3 = exp(pred3)
mean3 = mean((scale3-test3$y)^2)
mean3

boost4 = gbm(log(y) ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train4, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost4)
pred4 = predict(boost4, newdata=test2)
scale4 <- exp(pred4)
mean4 = mean((scale4-test2$y)^2)
mean4

boost5 = gbm(log(y) ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train5, distribution =  "gaussian", n.trees=3000, interaction.depth=3 )
summary(boost5)
```

```
pred5 = predict(boost5, newdata=test1)
scale5 <- exp(pred5)
mean5 = mean((scale5-test1$y)^2)
mean5

a = mean1 + mean2 + mean3 +  mean4 + mean5
a/5

#random forest
install.packages("randomForest")
library(randomForest)
rf = randomForest(y_yeo2 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + service_mode_2 + service_mode_3, data =
train, ntree = 100, importance =TRUE )
importance(rf)
rf

rf = randomForest(y_yeo2 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + service_mode_2 + service_mode_3, data =
train,ntree = 300,  importance =TRUE )
importance(rf)
rf
```

```
rf = randomForest(y_yeo2 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + service_mode_2 + service_mode_3, data =
train, ntree=500, importance =TRUE )
importance(rf)
rf

#assuming that the dataset in a randomic order i split it in 5 folds (17995/5=3599) and so create 5
train set
test1 = train[train$X<3601,]
test2 = train[3600<train$X & train$X<7201,]
test3 = train[7200<train$X & train$X<10801,]
test4 = train[10800<train$X & train$X<14401,]
test5 = train[14400<train$X,] #see than some a missing info
train1 = train[train$X<14401,] #without 5 fold
train2 = train[train$X<10801 | 14400<train$X,] #without 4 fold
train3 = train[train$X<7201 | 10800<train$X,] #without 3
train4 = train[train$X<3601 | 7200<train$X,] #without 2
train5 = train[train$X> 3600, ] #without1



#prediction
test <- read.csv("C:/Users/39346/OneDrive/Desktop/test_time.csv")
test <- dummy_cols(test, select_columns = "operator")
test <- dummy_cols(test, select_columns = "customer")
test <- dummy_cols(test, select_columns = "product")
test <- dummy_cols(test, select_columns = "work_group")
test<- dummy_cols(test, select_columns = "priority")
test<- dummy_cols(test, select_columns = "type")
test<- dummy_cols(test, select_columns = "channel")
test<- dummy_cols(test, select_columns = "service_center")
test <- dummy_cols(test, select_columns = "service_mode")
test <- dummy_cols(test, select_columns = "sla_contract")

pred = predict(boost, newdata=test)
install.packages("VGAM")
library("VGAM")
ciao<- yeo.johnson(pred, 0.06, inverse=TRUE)
```

```
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
hist(ciao)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_yeo09.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE,
col.names= FALSE)


#try with <10000
boost = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train_10000, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost) #channel2 day_year sla_contract service_center13 operator_2331 hour
work_group_121 service_center_121 service_center_8

boost = gbm(log(y) ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost) #channel2 day_year sla_contract service_center13 operator_2331 hour
work_group_121 service_center_121 service_center_8

pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.055, inverse=TRUE)
hist(ciao)
```

```
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost10000.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE,
col.names= FALSE)

#some final try
#yeo 0.9 total var level 3 tree 0.9
boost = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)

boost = gbm(y_yeo ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.06, inverse=TRUE)
hist(ciao)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_yeo06_level3.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE,
col.names= FALSE)
```

```
#shrinkage 0.001 livelli 4 tree 3000 yeo 0.9
boost = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=3000, interaction.depth=4)
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
hist(ciao)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_6.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)

#shrinkage 0.01 livelli 4 tree 3000 yeo 0.9
boost = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", shrinkage=0.01, n.trees=3000, interaction.depth=4)
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_7.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)
```

```
#shrinkage 0.01 livelli 4 tree 00 yeo 0.9
boost = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", shrinkage=0.01, n.trees=200, interaction.depth=4)
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_8.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)

#yeo 0.9 without month
boost = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + service_mode_2 + service_mode_3, data =
train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_9.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)
```

```
#yeo 0.9 without month with weekday
boost = gbm(y_yeo1 ~ month + weekday + day_year+ product_313 + product_203 + product_298 +
product_347 + product_457 + product_41 + product_12 + product_221 + product_158 + product_50 +
product_27 + product_679 + product_956 + product_168 + work_group_131 + work_group_35 +
work_group_29 + work_group_7 + work_group_31 + work_group_53 + work_group_27 +
work_group_26 + work_group_22 + work_group_121 + work_group_24 + work_group_50 +
work_group_20  + work_group_24  + work_group_33 + customer_5622 + customer_3752 +
customer_5709 + customer_3758 + customer_10467 + customer_3752 + customer_10853 +
customer_3032 + customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927
+ customer_888 + customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9
+type_17 + channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174
+ operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220
+ operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract + hour + service_mode_2 + service_mode_3, data =
train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_10.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)


#yeo 1 from 821
boost = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract +hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
```

```
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_11.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)


#yeo2
boost = gbm(y_yeo2 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
service_center_15 + forwarded + sla_contract +hour + month + service_mode_2 + service_mode_3,
data = train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.1, inverse=TRUE)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_13.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)


 #closing and open
train$opening <- as.factor(ifelse(train$hour>17 | train$hour<8 | train$hour == 13, 1, 0))
test$opening <- as.factor(ifelse(test$hour>17 | test$hour<8 | test$hour == 13, 1, 0))

boost = gbm(y_yeo1 ~ day_year+ product_313 + product_203 + product_298 + product_347 +
product_457 + product_41 + product_12 + product_221 + product_158 + product_50 + product_27 +
product_679 + product_956 + product_168 + work_group_131 + work_group_35 + work_group_29 +
work_group_7 + work_group_31 + work_group_53 + work_group_27 + work_group_26 +
work_group_22 + work_group_121 + work_group_24 + work_group_50 + work_group_20  +
work_group_24  + work_group_33 + customer_5622 + customer_3752 + customer_5709 +
customer_3758 + customer_10467 + customer_3752 + customer_10853 + customer_3032 +
customer_1+ customer_8912 + customer_771 + customer_10674 + customer_927 + customer_888 +
customer_918 + customer_3178 + priority + type_3 + type_5 + type_20 + type_9 +type_17 +
channel_2 + channel_5 + operator_155 + operator_911 + operator_3652 + operator_4174 +
operator_899 + operator_1941 + operator_2331+ operator_1213 + operator_1223 + operator_4220 +
operator_910 + operator_1223 +operator_1565 + operator_2331 + operator_2744 + operator_200 +
operator_2742 + operator_4198 + operator_2742 + operator_1218 + operator_4158 + operator_4828
+ operator_20 + operator_4804 + operator_225 + service_center_11 + service_center_13 +
service_center_8 + service_center_19 + service_center_1 + service_center_3 + service_center_7 +
```

```
service_center_15 + forwarded + sla_contract + opening + hour + month + service_mode_2 +
service_mode_3, data = train, distribution =  "gaussian", n.trees=5000, interaction.depth=3 )
summary(boost)
pred = predict(boost, newdata=test)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/boost_14.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)

#random forest 0,9 1000 alberi sqrt
install.packages("randomForest")
library(randomForest)
rf = randomForest(y_yeo1 ~ day_year + I(day_year^2) + opening + product_313 + product_203 +
product_298 + product_347 + product_457 + product_41 + product_12 + product_221 + product_158
+ product_50 + product_27 + product_679 + product_956 + product_168 + work_group_131 +
work_group_35 + work_group_29 + work_group_7 + work_group_31 + work_group_53 +
work_group_27 + work_group_26 + work_group_22 + work_group_121 + work_group_24 +
work_group_50 + work_group_20  + work_group_24  + work_group_33 + customer_5622 +
customer_3752 + customer_5709 + customer_3758 + customer_10467 + customer_3752 +
customer_10853 + customer_3032 + customer_1+ customer_8912 + customer_771 +
customer_10674 + customer_927 + customer_888 + customer_918 + customer_3178 + priority +
type_3 + type_5 + type_20 + type_9 +type_17 + channel_2 + channel_5 + operator_155 +
operator_911 + operator_3652 + operator_4174 + operator_899 + operator_1941 + operator_2331+
operator_1213 + operator_1223 + operator_4220 + operator_910 + operator_1223 +operator_1565 +
operator_2331 + operator_2744 + operator_200 + operator_2742 + operator_4198 + operator_2742 +
operator_1218 + operator_4158 + operator_4828 + operator_20 + operator_4804 + operator_225 +
service_center_11 + service_center_13 + service_center_8 + service_center_19 + service_center_1 +
service_center_3 + service_center_7 + service_center_15 + forwarded + sla_contract + hour + month
+ service_mode_2 + service_mode_3, data = train, ntree = 1000, mtry=10,  importance =TRUE )
importance(rf)
rf
pred = predict(rf, newdata=test)
ciao<- yeo.johnson(pred, 0.09, inverse=TRUE)
setwd("C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA CHALLENGE")
write.table(ciao, file="C:/Users/39346/OneDrive/Desktop/MACHINE LEARNING/FINAL DATA
CHALLENGE/rf3.txt", append = TRUE, sep = "\n", dec = ".", row.names = FALSE, col.names=
FALSE)
```