공공자전거수요예측\_빅개론텀프로젝트\_201823871\_박지헌

ParkJiHeon

2022 5 30

# 2022.05.07   
# 빅데이터개론 및 분석 프로젝트  
# e비즈니스학과 201823871 박지헌  
getwd()

## [1] "C:/Users/82104/my-workspace/R-workspace"

setwd('C:/Users/82104/my-workspace/R-workspace/')  
rm(list = ls())  
#install.packages('psych')  
library(psych)  
library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

library(doBy)  
library(randomForest)

## Warning: package 'randomForest' was built under R version 4.0.5

## randomForest 4.6-14

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

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':  
##   
## margin

## The following object is masked from 'package:psych':  
##   
## outlier

library(e1071)

## Warning: package 'e1071' was built under R version 4.0.5

library(rpart)

## Warning: package 'rpart' was built under R version 4.0.5

df = read.csv('2020.AI.bike-train.csv')  
  
str(df)

## 'data.frame': 1000 obs. of 11 variables:  
## $ id : int 3 6 7 8 9 13 14 16 19 20 ...  
## $ hour : int 20 13 6 23 18 2 3 21 9 14 ...  
## $ hour\_bef\_temperature : num 16.3 20.1 13.9 8.1 29.5 13.6 10.6 16 13.8 17.2 ...  
## $ hour\_bef\_precipitation: num 1 0 0 0 0 0 0 0 0 0 ...  
## $ hour\_bef\_windspeed : num 1.5 1.4 0.7 2.7 4.8 1.7 1.5 6 1.9 2.1 ...  
## $ hour\_bef\_humidity : num 89 48 79 54 7 80 58 21 64 32 ...  
## $ hour\_bef\_visibility : num 576 916 1382 946 2000 ...  
## $ hour\_bef\_ozone : num 0.027 0.042 0.033 0.04 0.057 0.027 0.038 0.05 0.039 0.025 ...  
## $ hour\_bef\_pm10 : num 76 73 32 75 27 34 62 90 93 64 ...  
## $ hour\_bef\_pm2.5 : num 33 40 19 64 11 15 33 28 19 19 ...  
## $ count : num 49 159 26 57 431 39 23 146 39 83 ...

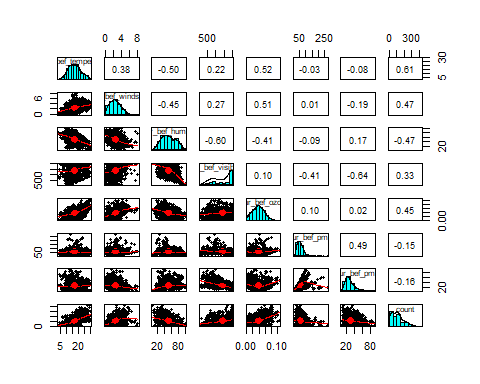
#table(is.na(df))  
#plot(df)  
df = df[-1:-2];head(df)

## hour\_bef\_temperature hour\_bef\_precipitation hour\_bef\_windspeed  
## 1 16.3 1 1.5  
## 2 20.1 0 1.4  
## 3 13.9 0 0.7  
## 4 8.1 0 2.7  
## 5 29.5 0 4.8  
## 6 13.6 0 1.7  
## hour\_bef\_humidity hour\_bef\_visibility hour\_bef\_ozone hour\_bef\_pm10  
## 1 89 576 0.027 76  
## 2 48 916 0.042 73  
## 3 79 1382 0.033 32  
## 4 54 946 0.040 75  
## 5 7 2000 0.057 27  
## 6 80 1073 0.027 34  
## hour\_bef\_pm2.5 count  
## 1 33 49  
## 2 40 159  
## 3 19 26  
## 4 64 57  
## 5 11 431  
## 6 15 39

df = df[-2];head(df)

## hour\_bef\_temperature hour\_bef\_windspeed hour\_bef\_humidity hour\_bef\_visibility  
## 1 16.3 1.5 89 576  
## 2 20.1 1.4 48 916  
## 3 13.9 0.7 79 1382  
## 4 8.1 2.7 54 946  
## 5 29.5 4.8 7 2000  
## 6 13.6 1.7 80 1073  
## hour\_bef\_ozone hour\_bef\_pm10 hour\_bef\_pm2.5 count  
## 1 0.027 76 33 49  
## 2 0.042 73 40 159  
## 3 0.033 32 19 26  
## 4 0.040 75 64 57  
## 5 0.057 27 11 431  
## 6 0.027 34 15 39

pairs.panels(df)



scale\_model <- caret::preProcess(df[,-8], method = 'range')  
df\_scaled <- predict(scale\_model, df)  
head(df\_scaled)

## hour\_bef\_temperature hour\_bef\_windspeed hour\_bef\_humidity hour\_bef\_visibility  
## 1 0.4907063 0.1875 0.8913043 0.2591051  
## 2 0.6319703 0.1750 0.4456522 0.4360042  
## 3 0.4014870 0.0875 0.7826087 0.6784599  
## 4 0.1858736 0.3375 0.5108696 0.4516129  
## 5 0.9814126 0.6000 0.0000000 1.0000000  
## 6 0.3903346 0.2125 0.7934783 0.5176899  
## hour\_bef\_ozone hour\_bef\_pm10 hour\_bef\_pm2.5 count  
## 1 0.2330097 0.24902724 0.29629630 49  
## 2 0.3786408 0.23735409 0.38271605 159  
## 3 0.2912621 0.07782101 0.12345679 26  
## 4 0.3592233 0.24513619 0.67901235 57  
## 5 0.5242718 0.05836576 0.02469136 431  
## 6 0.2330097 0.08560311 0.07407407 39

set.seed(123)  
idx\_train = sample(1:nrow(df\_scaled), size = 800)  
df\_train = df\_scaled[idx\_train, ]  
df\_test = df\_scaled[-idx\_train, ]  
  
  
#https://bioinformaticsandme.tistory.com/290  
# 다중회귀분석  
m1 = lm(count~., data = df\_train)  
summary(m1)

##   
## Call:  
## lm(formula = count ~ ., data = df\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -179.427 -42.168 -6.633 36.933 201.904   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.554 24.473 -0.145 0.88458   
## hour\_bef\_temperature 180.032 14.368 12.530 < 2e-16 \*\*\*  
## hour\_bef\_windspeed 99.354 15.603 6.368 3.25e-10 \*\*\*  
## hour\_bef\_humidity -47.653 16.988 -2.805 0.00515 \*\*   
## hour\_bef\_visibility 18.988 13.624 1.394 0.16379   
## hour\_bef\_ozone 48.947 15.032 3.256 0.00118 \*\*   
## hour\_bef\_pm10 -111.079 22.571 -4.921 1.05e-06 \*\*\*  
## hour\_bef\_pm2.5 27.961 17.042 1.641 0.10125   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 60.95 on 792 degrees of freedom  
## Multiple R-squared: 0.493, Adjusted R-squared: 0.4885   
## F-statistic: 110 on 7 and 792 DF, p-value: < 2.2e-16

p1 = predict(m1, df\_test);cor(df\_test$count, p1)

## [1] 0.6549888

m1 = update(m1, .~.-hour\_bef\_visibility-hour\_bef\_pm2.5)  
summary(m1)

##   
## Call:  
## lm(formula = count ~ hour\_bef\_temperature + hour\_bef\_windspeed +   
## hour\_bef\_humidity + hour\_bef\_ozone + hour\_bef\_pm10, data = df\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -178.531 -42.936 -6.772 37.143 205.557   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 26.81 12.55 2.137 0.03292 \*   
## hour\_bef\_temperature 177.36 14.21 12.481 < 2e-16 \*\*\*  
## hour\_bef\_windspeed 95.86 15.35 6.246 6.87e-10 \*\*\*  
## hour\_bef\_humidity -61.91 11.98 -5.168 2.99e-07 \*\*\*  
## hour\_bef\_ozone 48.63 14.91 3.262 0.00115 \*\*   
## hour\_bef\_pm10 -112.82 17.83 -6.328 4.14e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 60.99 on 794 degrees of freedom  
## Multiple R-squared: 0.4911, Adjusted R-squared: 0.4879   
## F-statistic: 153.2 on 5 and 794 DF, p-value: < 2.2e-16

p1 = predict(m1, df\_test);cor(df\_test$count, p1)

## [1] 0.6492942

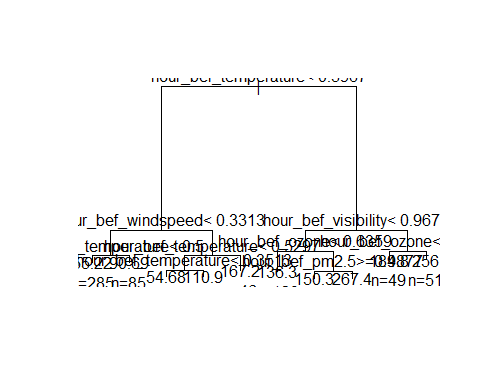
# 의사결정나무  
m2 = rpart(count~., data = df\_train)  
summary(m2)

## Call:  
## rpart(formula = count ~ ., data = df\_train)  
## n= 800   
##   
## CP nsplit rel error xerror xstd  
## 1 0.31840259 0 1.0000000 1.0013265 0.05338991  
## 2 0.05600917 1 0.6815974 0.7025181 0.03751870  
## 3 0.04611906 2 0.6255882 0.6549086 0.03511286  
## 4 0.04412372 3 0.5794692 0.6527803 0.03607128  
## 5 0.02844296 4 0.5353455 0.6138085 0.03301570  
## 6 0.01889831 5 0.5069025 0.5767824 0.03005437  
## 7 0.01546257 6 0.4880042 0.5818669 0.03043873  
## 8 0.01419748 7 0.4725416 0.5741271 0.03074476  
## 9 0.01281403 8 0.4583441 0.5688961 0.03106418  
## 10 0.01000000 9 0.4455301 0.5532755 0.02993332  
##   
## Variable importance  
## hour\_bef\_temperature hour\_bef\_ozone hour\_bef\_windspeed   
## 46 19 10   
## hour\_bef\_humidity hour\_bef\_visibility hour\_bef\_pm2.5   
## 8 7 5   
## hour\_bef\_pm10   
## 5   
##   
## Node number 1: 800 observations, complexity param=0.3184026  
## mean=113.1, MSE=7253.677   
## left son=2 (544 obs) right son=3 (256 obs)  
## Primary splits:  
## hour\_bef\_temperature < 0.5966543 to the left, improve=0.3184026, (0 missing)  
## hour\_bef\_ozone < 0.4029126 to the left, improve=0.2058068, (0 missing)  
## hour\_bef\_windspeed < 0.31875 to the left, improve=0.2043652, (0 missing)  
## hour\_bef\_humidity < 0.375 to the right, improve=0.1629305, (0 missing)  
## hour\_bef\_visibility < 0.8231009 to the left, improve=0.1027314, (0 missing)  
## Surrogate splits:  
## hour\_bef\_ozone < 0.4708738 to the left, agree=0.771, adj=0.285, (0 split)  
## hour\_bef\_humidity < 0.298913 to the right, agree=0.714, adj=0.105, (0 split)  
## hour\_bef\_windspeed < 0.41875 to the left, agree=0.701, adj=0.066, (0 split)  
## hour\_bef\_pm10 < 0.8249027 to the left, agree=0.681, adj=0.004, (0 split)  
##   
## Node number 2: 544 observations, complexity param=0.05600917  
## mean=80.13235, MSE=4123.28   
## left son=4 (370 obs) right son=5 (174 obs)  
## Primary splits:  
## hour\_bef\_windspeed < 0.33125 to the left, improve=0.14489910, (0 missing)  
## hour\_bef\_temperature < 0.5 to the left, improve=0.12207470, (0 missing)  
## hour\_bef\_humidity < 0.6902174 to the right, improve=0.10821430, (0 missing)  
## hour\_bef\_ozone < 0.3932039 to the left, improve=0.06505326, (0 missing)  
## hour\_bef\_visibility < 0.8311655 to the left, improve=0.05511129, (0 missing)  
## Surrogate splits:  
## hour\_bef\_ozone < 0.3737864 to the left, agree=0.739, adj=0.184, (0 split)  
## hour\_bef\_humidity < 0.2880435 to the right, agree=0.715, adj=0.109, (0 split)  
## hour\_bef\_pm2.5 < 0.06790123 to the right, agree=0.695, adj=0.046, (0 split)  
## hour\_bef\_temperature < 0.5780669 to the left, agree=0.686, adj=0.017, (0 split)  
## hour\_bef\_visibility < 0.05931322 to the right, agree=0.686, adj=0.017, (0 split)  
##   
## Node number 3: 256 observations, complexity param=0.04611906  
## mean=183.1562, MSE=6688.304   
## left son=6 (156 obs) right son=7 (100 obs)  
## Primary splits:  
## hour\_bef\_visibility < 0.9677419 to the left, improve=0.15630490, (0 missing)  
## hour\_bef\_windspeed < 0.30625 to the left, improve=0.14476000, (0 missing)  
## hour\_bef\_ozone < 0.4417476 to the left, improve=0.13417880, (0 missing)  
## hour\_bef\_pm2.5 < 0.1049383 to the right, improve=0.08542021, (0 missing)  
## hour\_bef\_temperature < 0.8048327 to the left, improve=0.08241376, (0 missing)  
## Surrogate splits:  
## hour\_bef\_pm2.5 < 0.1790123 to the right, agree=0.820, adj=0.54, (0 split)  
## hour\_bef\_pm10 < 0.118677 to the right, agree=0.789, adj=0.46, (0 split)  
## hour\_bef\_humidity < 0.2663043 to the right, agree=0.707, adj=0.25, (0 split)  
## hour\_bef\_windspeed < 0.46875 to the left, agree=0.660, adj=0.13, (0 split)  
## hour\_bef\_ozone < 0.0631068 to the right, agree=0.629, adj=0.05, (0 split)  
##   
## Node number 4: 370 observations, complexity param=0.01419748  
## mean=63.37027, MSE=2918.909   
## left son=8 (285 obs) right son=9 (85 obs)  
## Primary splits:  
## hour\_bef\_temperature < 0.5 to the left, improve=0.07628466, (0 missing)  
## hour\_bef\_humidity < 0.6902174 to the right, improve=0.06543619, (0 missing)  
## hour\_bef\_windspeed < 0.16875 to the left, improve=0.04852907, (0 missing)  
## hour\_bef\_pm10 < 0.1575875 to the right, improve=0.02588153, (0 missing)  
## hour\_bef\_visibility < 0.237513 to the left, improve=0.02105023, (0 missing)  
## Surrogate splits:  
## hour\_bef\_pm10 < 0.4494163 to the left, agree=0.776, adj=0.024, (0 split)  
## hour\_bef\_ozone < 0.004854369 to the right, agree=0.773, adj=0.012, (0 split)  
##   
## Node number 5: 174 observations, complexity param=0.02844296  
## mean=115.7759, MSE=4816.381   
## left son=10 (128 obs) right son=11 (46 obs)  
## Primary splits:  
## hour\_bef\_temperature < 0.5297398 to the left, improve=0.19694860, (0 missing)  
## hour\_bef\_humidity < 0.7228261 to the right, improve=0.16592730, (0 missing)  
## hour\_bef\_visibility < 0.194589 to the left, improve=0.12202270, (0 missing)  
## hour\_bef\_windspeed < 0.50625 to the left, improve=0.08573862, (0 missing)  
## hour\_bef\_ozone < 0.3932039 to the left, improve=0.08430550, (0 missing)  
## Surrogate splits:  
## hour\_bef\_ozone < 0.6165049 to the left, agree=0.747, adj=0.043, (0 split)  
## hour\_bef\_humidity < 0.201087 to the right, agree=0.741, adj=0.022, (0 split)  
## hour\_bef\_pm10 < 0.5252918 to the left, agree=0.741, adj=0.022, (0 split)  
##   
## Node number 6: 156 observations, complexity param=0.04412372  
## mean=157.2692, MSE=5323.094   
## left son=12 (123 obs) right son=13 (33 obs)  
## Primary splits:  
## hour\_bef\_ozone < 0.6359223 to the left, improve=0.30834130, (0 missing)  
## hour\_bef\_temperature < 0.7527881 to the left, improve=0.14752400, (0 missing)  
## hour\_bef\_pm2.5 < 0.2777778 to the left, improve=0.09612175, (0 missing)  
## hour\_bef\_windspeed < 0.30625 to the left, improve=0.09081566, (0 missing)  
## hour\_bef\_pm10 < 0.1498054 to the left, improve=0.06163219, (0 missing)  
## Surrogate splits:  
## hour\_bef\_temperature < 0.8698885 to the left, agree=0.846, adj=0.273, (0 split)  
## hour\_bef\_pm2.5 < 0.5617284 to the left, agree=0.808, adj=0.091, (0 split)  
## hour\_bef\_humidity < 0.1684783 to the right, agree=0.801, adj=0.061, (0 split)  
## hour\_bef\_pm10 < 0.8054475 to the left, agree=0.801, adj=0.061, (0 split)  
##   
## Node number 7: 100 observations, complexity param=0.01889831  
## mean=223.54, MSE=6141.768   
## left son=14 (49 obs) right son=15 (51 obs)  
## Primary splits:  
## hour\_bef\_ozone < 0.4708738 to the left, improve=0.17855730, (0 missing)  
## hour\_bef\_windspeed < 0.33125 to the left, improve=0.16171340, (0 missing)  
## hour\_bef\_temperature < 0.6263941 to the left, improve=0.12345950, (0 missing)  
## hour\_bef\_humidity < 0.375 to the right, improve=0.07605974, (0 missing)  
## hour\_bef\_pm2.5 < 0.04320988 to the right, improve=0.06371920, (0 missing)  
## Surrogate splits:  
## hour\_bef\_humidity < 0.3097826 to the right, agree=0.75, adj=0.490, (0 split)  
## hour\_bef\_windspeed < 0.38125 to the left, agree=0.72, adj=0.429, (0 split)  
## hour\_bef\_temperature < 0.7843866 to the left, agree=0.69, adj=0.367, (0 split)  
## hour\_bef\_pm10 < 0.118677 to the left, agree=0.60, adj=0.184, (0 split)  
## hour\_bef\_visibility < 0.9921956 to the right, agree=0.59, adj=0.163, (0 split)  
##   
## Node number 8: 285 observations  
## mean=55.22105, MSE=2483.899   
##   
## Node number 9: 85 observations  
## mean=90.69412, MSE=3408.212   
##   
## Node number 10: 128 observations, complexity param=0.01281403  
## mean=97.3125, MSE=3564.934   
## left son=20 (31 obs) right son=21 (97 obs)  
## Primary splits:  
## hour\_bef\_temperature < 0.3513011 to the left, improve=0.16295690, (0 missing)  
## hour\_bef\_visibility < 0.1938085 to the left, improve=0.14414290, (0 missing)  
## hour\_bef\_humidity < 0.7228261 to the right, improve=0.13292210, (0 missing)  
## hour\_bef\_ozone < 0.4320388 to the left, improve=0.10938950, (0 missing)  
## hour\_bef\_windspeed < 0.51875 to the left, improve=0.04397022, (0 missing)  
## Surrogate splits:  
## hour\_bef\_humidity < 0.8423913 to the right, agree=0.828, adj=0.290, (0 split)  
## hour\_bef\_visibility < 0.1545265 to the left, agree=0.820, adj=0.258, (0 split)  
## hour\_bef\_ozone < 0.1019417 to the left, agree=0.773, adj=0.065, (0 split)  
## hour\_bef\_pm10 < 0.348249 to the right, agree=0.766, adj=0.032, (0 split)  
##   
## Node number 11: 46 observations  
## mean=167.1522, MSE=4710.564   
##   
## Node number 12: 123 observations  
## mean=136.2846, MSE=2762.529   
##   
## Node number 13: 33 observations, complexity param=0.01546257  
## mean=235.4848, MSE=7108.007   
## left son=26 (9 obs) right son=27 (24 obs)  
## Primary splits:  
## hour\_bef\_pm2.5 < 0.4876543 to the right, improve=0.3825323, (0 missing)  
## hour\_bef\_pm10 < 0.2490272 to the right, improve=0.2335694, (0 missing)  
## hour\_bef\_visibility < 0.498179 to the left, improve=0.2266052, (0 missing)  
## hour\_bef\_windspeed < 0.35625 to the right, improve=0.1724429, (0 missing)  
## hour\_bef\_temperature < 0.7509294 to the left, improve=0.1181770, (0 missing)  
## Surrogate splits:  
## hour\_bef\_pm10 < 0.2490272 to the right, agree=0.909, adj=0.667, (0 split)  
## hour\_bef\_visibility < 0.3959417 to the left, agree=0.879, adj=0.556, (0 split)  
## hour\_bef\_temperature < 0.7509294 to the left, agree=0.848, adj=0.444, (0 split)  
## hour\_bef\_ozone < 0.6504854 to the left, agree=0.758, adj=0.111, (0 split)  
##   
## Node number 14: 49 observations  
## mean=189.7551, MSE=4746.226   
##   
## Node number 15: 51 observations  
## mean=256, MSE=5332.275   
##   
## Node number 20: 31 observations  
## mean=54.67742, MSE=2359.315   
##   
## Node number 21: 97 observations  
## mean=110.9381, MSE=3183.646   
##   
## Node number 26: 9 observations  
## mean=150.3333, MSE=3886.889   
##   
## Node number 27: 24 observations  
## mean=267.4167, MSE=4577.243

p2 = predict(m2, df\_test);cor(df\_test$count, p2)

## [1] 0.6707049

plot(m2)  
text(m2, use.n = TRUE)



# 랜덤 포레스트트  
for(nt in c(20,50,100,200,300,400,500)) {  
 m = randomForest(count~., data = df\_train, ntree = nt)  
 summary(m)  
 p = predict(m, df\_train)  
 cat("ntree =",nt,"acc =",cor(df\_train$count, p),"\n")  
}

## ntree = 20 acc = 0.9568524   
## ntree = 50 acc = 0.960718   
## ntree = 100 acc = 0.9631481   
## ntree = 200 acc = 0.965282   
## ntree = 300 acc = 0.9650979   
## ntree = 400 acc = 0.9649653   
## ntree = 500 acc = 0.9646034

for(nd in 1:30) {  
 m = randomForest(count~., data = df\_train, maxnodes=nd)  
 summary(m)  
 p = predict(m, df\_train)  
 cat("node =", nd, "acc =",cor(df\_train$count, p),"\n")  
}

## Warning in cor(df\_train$count, p): 표준편차가 0입니다

## node = 1 acc = NA   
## node = 2 acc = 0.6960168   
## node = 3 acc = 0.7027021   
## node = 4 acc = 0.731792   
## node = 5 acc = 0.7357146   
## node = 6 acc = 0.7418679   
## node = 7 acc = 0.751885   
## node = 8 acc = 0.7609687   
## node = 9 acc = 0.7632106   
## node = 10 acc = 0.7671117   
## node = 11 acc = 0.7704909   
## node = 12 acc = 0.7746093   
## node = 13 acc = 0.7777949   
## node = 14 acc = 0.7829474   
## node = 15 acc = 0.7874892   
## node = 16 acc = 0.7914787   
## node = 17 acc = 0.7943977   
## node = 18 acc = 0.795839   
## node = 19 acc = 0.799394   
## node = 20 acc = 0.8007481   
## node = 21 acc = 0.8035704   
## node = 22 acc = 0.8061884   
## node = 23 acc = 0.8083035   
## node = 24 acc = 0.809971   
## node = 25 acc = 0.8135564   
## node = 26 acc = 0.8164567   
## node = 27 acc = 0.8178764   
## node = 28 acc = 0.8218455   
## node = 29 acc = 0.8235538   
## node = 30 acc = 0.8269893

OPT\_NT = 50   
OPT\_ND = 29   
  
m3 = randomForest(count~., data = df\_train, ntree = OPT\_NT, maxnodes=OPT\_ND)  
summary(m3)

## Length Class Mode   
## call 5 -none- call   
## type 1 -none- character  
## predicted 800 -none- numeric   
## mse 50 -none- numeric   
## rsq 50 -none- numeric   
## oob.times 800 -none- numeric   
## importance 7 -none- numeric   
## importanceSD 0 -none- NULL   
## localImportance 0 -none- NULL   
## proximity 0 -none- NULL   
## ntree 1 -none- numeric   
## mtry 1 -none- numeric   
## forest 11 -none- list   
## coefs 0 -none- NULL   
## y 800 -none- numeric   
## test 0 -none- NULL   
## inbag 0 -none- NULL   
## terms 3 terms call

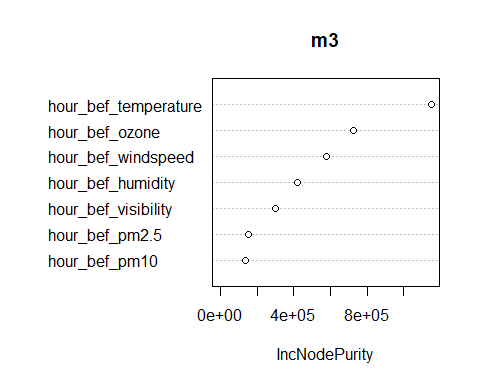
p3 = predict(m3, df\_test);cor(df\_test$count, p3)

## [1] 0.7063283

varUsed(m3)

## [1] 267 209 219 166 208 177 154

varImpPlot(m3)



# SVM  
m4 = svm(count~., data = df\_train)  
summary(m4)

##   
## Call:  
## svm(formula = count ~ ., data = df\_train)  
##   
##   
## Parameters:  
## SVM-Type: eps-regression   
## SVM-Kernel: radial   
## cost: 1   
## gamma: 0.1428571   
## epsilon: 0.1   
##   
##   
## Number of Support Vectors: 669

p4 = predict(m4, df\_test);cor(df\_test$count, p4)

## [1] 0.731444

#cv  
control = trainControl(method = 'cv', number=10)  
formular = count~.  
  
s = train(formular, data=df\_train, method='svmLinear', metric='RMSE',trControl=control)  
sp = train(formular, data=df\_train, method='svmPoly', metric='RMSE',trControl=control)  
sr = train(formular, data=df\_train, method='svmRadial', metric='RMSE',trControl=control)  
rf = train(formular, data=df\_train, method='rf', metric='RMSE',trControl=control)  
r = train(formular, data=df\_train, method='rpart', metric='RMSE',trControl=control)

## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :  
## There were missing values in resampled performance measures.

lm = train(formular, data=df\_train, method='lm', metric='RMSE',trControl=control)  
k = train(formular, data=df\_train, method='knn', metric='RMSE',trControl=control)  
  
resamp = resamples(list(svm=s, svmPoly=sp, svmRadial=sr, RF=rf, decisionTree=r, regression=lm,kNN=k))  
summary(resamp)

##   
## Call:  
## summary.resamples(object = resamp)  
##   
## Models: svm, svmPoly, svmRadial, RF, decisionTree, regression, kNN   
## Number of resamples: 10   
##   
## MAE   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## svm 43.13786 45.35429 48.11822 48.13678 51.16942 53.57101 0  
## svmPoly 35.14732 43.30141 43.96527 43.61946 45.83450 46.85753 0  
## svmRadial 36.94292 38.69669 43.54428 43.11506 46.93010 49.21813 0  
## RF 34.13239 41.04898 42.69341 42.72032 45.39393 49.53977 0  
## decisionTree 50.81316 52.24391 52.96775 53.79718 55.31132 58.75406 0  
## regression 40.42182 47.59823 48.61142 48.40426 51.59903 52.60707 0  
## kNN 35.44040 42.45309 45.44154 44.72864 49.06741 50.80651 0  
##   
## RMSE   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## svm 54.08844 58.94651 62.53951 61.64297 65.25336 67.14099 0  
## svmPoly 44.08269 57.32766 59.03481 58.26050 61.08627 64.48954 0  
## svmRadial 49.84325 51.21786 55.35724 57.01574 63.80834 66.15885 0  
## RF 43.86622 52.11569 56.57914 55.23679 59.01678 61.48157 0  
## decisionTree 64.03444 65.25407 67.60524 67.94019 70.84050 72.34908 0  
## regression 50.58360 59.17334 59.97731 61.25020 65.48234 68.59700 0  
## kNN 45.82908 53.83628 58.74979 57.04485 61.96956 63.95826 0  
##   
## Rsquared   
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
## svm 0.3190197 0.4597288 0.5030335 0.4945816 0.5493189 0.6391917 0  
## svmPoly 0.4166440 0.4908605 0.5383102 0.5487049 0.5981006 0.7129942 0  
## svmRadial 0.4351235 0.4692061 0.6056482 0.5657401 0.6239401 0.6838168 0  
## RF 0.4712694 0.5216273 0.5892842 0.5826921 0.6227217 0.7186424 0  
## decisionTree 0.2160676 0.3420881 0.4001504 0.3766509 0.4240618 0.4665680 0  
## regression 0.3662291 0.3947241 0.4993862 0.4823224 0.5596922 0.5913825 0  
## kNN 0.4359397 0.4835567 0.5368156 0.5529318 0.6267695 0.6932060 0

sort(resamp,decreasing = TRUE)

## [1] "decisionTree" "regression" "svm" "kNN" "svmPoly"   
## [6] "svmRadial" "RF"

dotplot(resamp)

