DATA 624: Project 2

DATA 624 - Predictive Analytics

Group 2

Group Members:

Vinicio Haro

Sang Yoon (Andy) Hwang

Julian McEachern

Jeremy O'Brien

Bethany Poulin

10 December 2019

Contents

File for final submission of Project 2.

1   Model Performance

- Set1 = Caret: bagImputed; no additional pre-processing

- Set2 = Caret: bagImputed; PreP method=c('center', 'scale', 'nzv', 'BoxCox')

Train Performance:

Table 1.1: Train1 Performance

MAPE	RMSE	RSquared	MAE	Method
1.0948	0.5916	0.6757	0.4319	rf
1.1629	0.5500	0.6979	0.3938	cubist
1.2664	0.6985	0.5165	0.5037	svmRadial
1.5170	0.6921	0.5261	0.5153	earth

Table 1.2: Train2 Performance

MAPE	RMSE	RSquared	MAE	Method
0.0081	0.0965	0.6904	0.0688	cubist
0.0088	0.1030	0.6735	0.0751	rf
0.0104	0.1224	0.5042	0.0883	svmRadial
0.0112	0.1265	0.4690	0.0954	earth

Test Accuracy:

Table 1.3: Test1 Performance

MAPE	RMSE	Rsquared	MAE	Method
0.4930	0.2234	0.9539	0.1578	cubist
0.5351	0.3265	0.9568	0.2466	rf
0.9987	0.6076	0.6385	0.4101	svmRadial
1.6688	0.7260	0.5226	0.5269	earth

Table 1.4: Test2 Performance

MAPE	RMSE	Rsquared	MAE	Method
0.0029	0.0367	0.9586	0.0250	cubist
0.0037	0.0436	0.9596	0.0311	rf
0.0084	0.1064	0.6307	0.0717	svmRadial
0.0107	0.1193	0.5257	0.0908	earth

2   New Model : RPART

Cool graph if we add this or replace RF with it –>

Metrics

Train with transformation:

```r
library(party)
library(partykit)
library(caret)
library(MLmetrics)
# install.packages('rattle')
library(rattle)
set.seed(58677)

grid_rpart <- expand.grid(maxdepth = 1:20)
fit.rpart <- train(PH ~ ., data = train_trans, metric = "RMSE",
 method = "rpart2", tuneGrid = grid_rpart, tuneLength = tl,
 trControl = trC)

rp.PERF <- rbind(getTrainPerf(fit.rpart))
rp.MAPE <- cbind(MAPE = MAPE(fit.rpart$pred$pred, fit.rpart$pred$obs))
rp.ACC <- cbind(rp.MAPE, rp.PERF)




fit.rpart2 <- train(PH ~ ., data = train_trans, metric = "RMSE",
 method = "rpart", tuneLength = tl, trControl = trC)

rp.PERF2 <- rbind(getTrainPerf(fit.rpart2))
rp.MAPE2 <- cbind(MAPE = MAPE(fit.rpart2$pred$pred, fit.rpart2$pred$obs))
rp.ACC2 <- cbind(rp.MAPE2, rp.PERF2)


rbind(rp.ACC, rp.ACC2) %>% kable() %>% kable_styling()
```
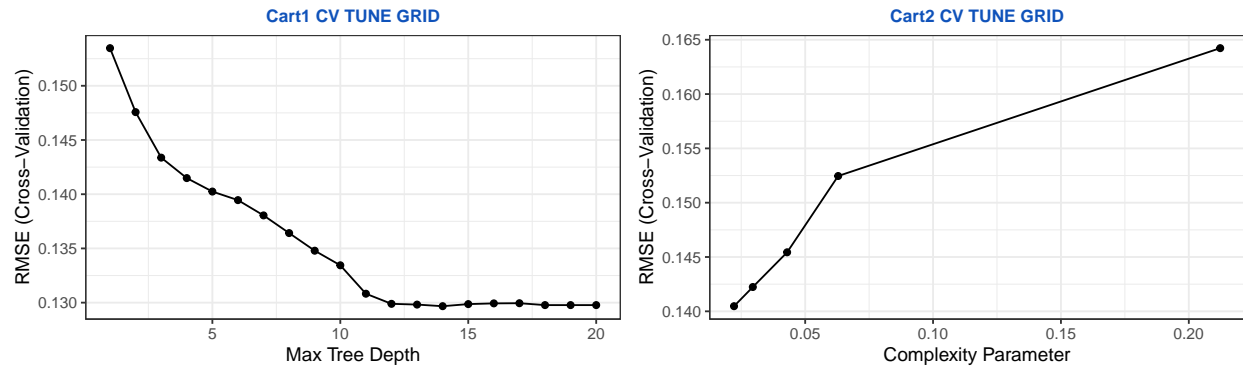
MAPE	TrainRMSE	TrainRsquared	TrainMAE	method
0.0116967	0.1296730	0.4422622	0.0996211	rpart2
0.0129475	0.1404728	0.3433799	0.1102129	rpart

Tune Grid

Cool Plot

Visualization of Cart1 Trees Influence on pH

Visualization of Cart2 Trees Influence on pH
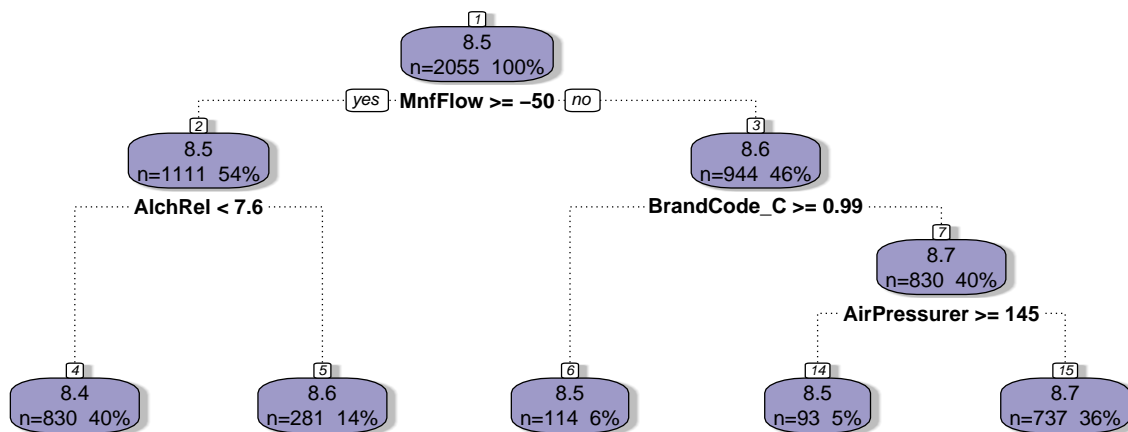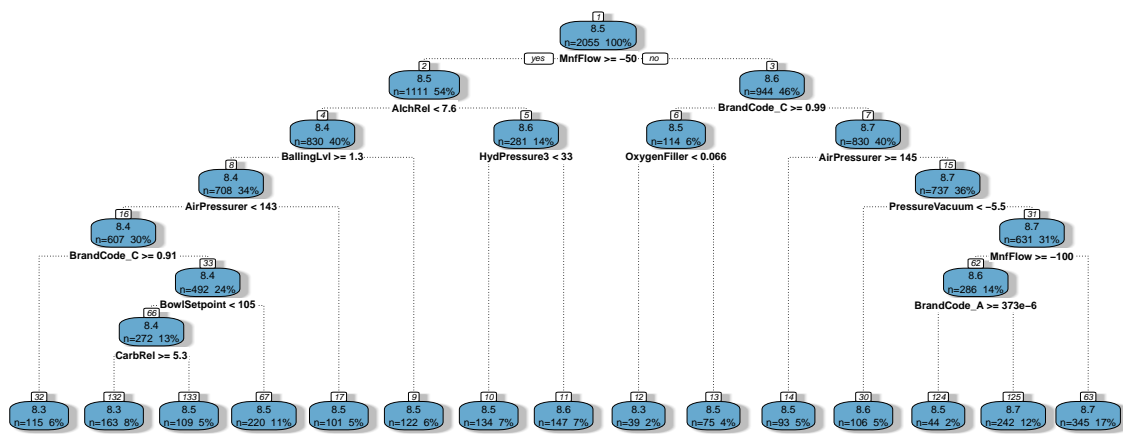
ALT PLOT

Do not recomment. Seems much more difficult to customize.

MnfFlow

2 ≥ −49.9   < −49.9   17

AlchRel   BrandCode_C

3   14   18   21

BallingLvl   HvdPressure   OxygenFiller   AirPressurer

4   13   15   16   19   20   22   23

AirPressurer
12
terminal
node

5   < 143.3

BrandCode_C   PressureVacuum   25

6   13   < 0.9   24   MnfFlow   29
terminal   terminal
node   7   node   26   terminal
                                            node
BowlSetpoint
8   11   BrandCode_A
        terminal   27   28
CarbRel   node   terminal terminal
9   10   node   node
terminal terminal
node   node

13 15 16 19 20
terminal terminal terminal terminal terminal
node   node   node   node   node

1

MnfFlow

≥ −49.9   < −49.9

2   5

AlchRel   BrandCode_C

< 7.56   ≥ 7.56   ≥ 0.994   < 0.994

3   4   6   7

terminal   terminal   terminal   AirPressurer
node   node   node

                            ≥ 144.7   < 144.7

                        8   9
                    terminal   terminal
                    node   node