To Churn or Not to Churn

REPORT 3

Aretha Kassegnin | Explorations in Data Science | 12-9-22

Part 1

To establish a bassline for this data I concluded that about 86% of customers will not churn. This is because based on the data there were more people didn't churn then churn to confirm this, I created a CART model to double check that I was right.

```
1) root 2509 356 False. (0.85811080 0.14188920)
  2) Day.Mins< 264.5 2361 264 False. (0.88818297 0.11181703)
    4) CustServ.Calls< 3.5 2172 166 False. (0.92357274 0.07642726) *
    5) CustServ.Calls>=3.5 189  91 True. (0.48148148 0.51851852)
     10) Day.Mins>=173.55 85  13 False. (0.84705882 0.15294118) *
     11) Day.Mins< 173.55 104  19 True. (0.18269231 0.81730769) *
  3) Day.Mins>=264.5 148  56 True. (0.37837838 0.62162162)
    6) VMail.Plan=yes 35   4 False. (0.88571429 0.11428571) *
    7) VMail.Plan=no 113  25 True. (0.22123894 0.77876106) *
```

Figure 1-1 root gives us a baseline model

Here below is the C5.0 model for predicting customer's churn

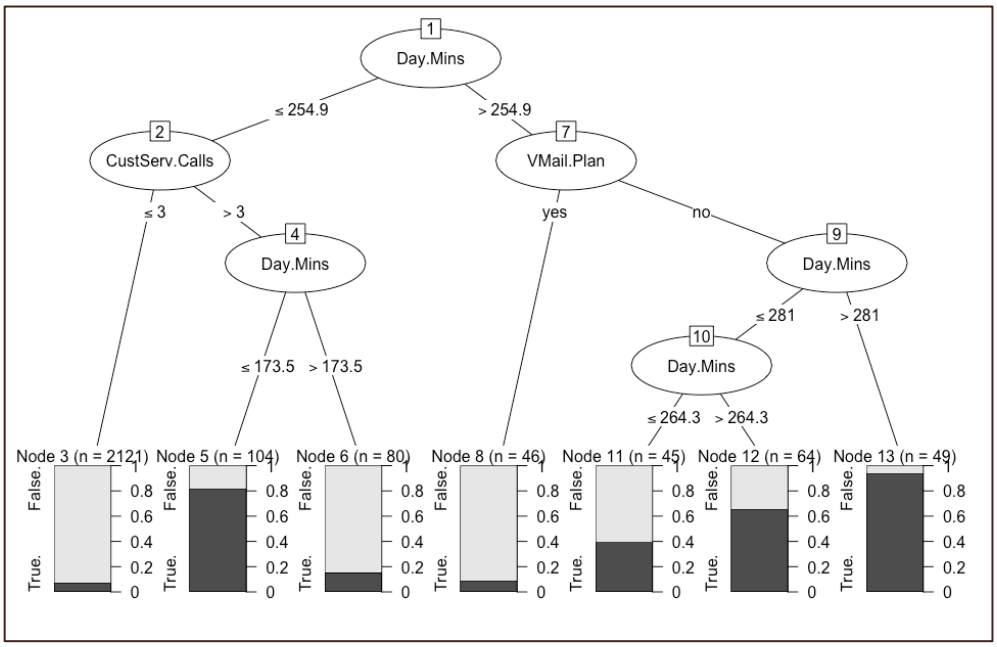

Figure 1-1 C5.0 model predicting customer's churn

The three decision rules that apply to the largest segment of the data as shown in Figure 1-2 is the variables Day.Mins (How much the customer uses their phone during the day), CustServ.Calls (The number of calls the customer makes to Customer Service) and VMail.Plan(Whether the customer has the Voicemail Plan) .

PART 2

	Predicted	
Actual	Churn False	Churn True
Churn False	676	21
Churn True	75	52

To find accuracy its (676+52)/824= .883 = 88.3%. This isn't so good because its below my baseline accuracy that I thought was prefect because of the. CART model this means that my C5 model isn't prefect.

The sensitivity for this data is about 52/(75+52)= .44= 44% This is not good because it needs to be close to 100%. But it shows that the data isn't just picking people that didn't churn sometimes people do churn.

The specificity of this model is 21/(676+21)= 3.01% This is good because it's a low score and it is not close to 100% with means we are not classifying everyone as not churned.

The C5.0 Model is good predicted accuracy my bassline was 86% and my model was 88.3% which is really good but it's not so great with sensitivity and specificity which is where the model fails.

Appendix

```
#Report 3

1. #Baseline

summary(churn_test$Churn.)

summary(churn_train$Churn.)


#based on the training data about 85% of people will not churn

#2

library(caret)
```

```
library(C50)
mod1 <- C5.0(Churn.~.,data = churn_train)
summary(mod1)
plot(mod1)

mod1.test <- predict(object = mod1, newdata = churn_test)
table(churn_test$Churn.,mod1.test)
summary(mod1.test)
dim(churn_test)[1]

library(rpart)
cart01 <- rpart(Churn. ~ .,
                data = churn_train,
                method = "class")
library(rpart.plot)
rpart.plot(cart01, type = 4, extra = 102)
dim(churn_test)[1]
```