Telecom Machine Learning

Nirmal Patel

November 13, 2018

Machine Learning Questions

How do you frame your main question as a machine learning problem?

My telecom churn analysis question can be frames as a machine learning question such as which factors are most influencial in customer churning. This would allow me to run machine learning models such as logistic regression as a binomial to figure out if a customer churns or not.

Is it a supervised or unsupervised problem?

The dataset I have is more of a supervised problem because the data is given in a well structured with varous input and output variables. in this case the output variable we are testing it the churning of a customer.

If it is supervised, is it a regression or a classification?

My problem is a regression problem due to churning being a binary variable with two factors such as yes or no. Though if we wanted to figure out which contract type a person would be more likely to choose then it would be classification.

What are the main features (also called independent variables or predictors) that you'll use?

Some of the independent variables include gender, SeniorCitizen, tenure, MultipleLines, InternetService, Contract, PaperlessBilling, PaymentMethod, and MonthlyCharges. The Dependent Variable is Churn of the customer.

Which machine learning technique will you use?

I will use logistic learning for figuring out churning of the customers.

How will you evaluate the success of your machine learning technique? What metric will you use?

The success of my model will be evaluated by the accuracy of logistic regression.

Machine Learning Models and Analysis

Logistic Regression

```
churnmodel <-
glm(Churn~gender+SeniorCitizen+Partner+Dependents+tenure+PhoneService+Multipl
eLines+InternetService+OnlineSecurity+OnlineBackup+DeviceProtection+TechSuppo
rt+StreamingTV+StreamingMovies+Contract+PaperlessBilling+PaymentMethod+Monthl
yCharges,data=telecom, family="binomial")
summary(churnmodel)

##
## Call:
## glm(formula = Churn ~ gender + SeniorCitizen + Partner + Dependents +
##     tenure + PhoneService + MultipleLines + InternetService +
##     OnlineSecurity + OnlineBackup + DeviceProtection + TechSupport +
##     StreamingTV + StreamingMovies + Contract + PaperlessBilling +
##     PaymentMethod + MonthlyCharges, family = "binomial", data = telecom)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9780  -0.6707  -0.2946   0.6918   3.1454
##
## Coefficients: (7 not defined because of singularities)
##                                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             0.612080   0.811986   0.754  0.45097
## genderMale                             -0.020514   0.064885  -0.316  0.75189
## SeniorCitizen                           0.217015   0.084920   2.556  0.01060
## PartnerYes                             -0.002440   0.077741  -0.031  0.97496
## DependentsYes                          -0.167071   0.089678  -1.863  0.06246
## tenure                                 -0.034172   0.002366 -14.443  < 2e-16
## PhoneServiceYes                         0.165499   0.652460   0.254  0.79976
## MultipleLinesNo phone service                 NA         NA      NA       NA
## MultipleLinesYes                        0.462796   0.178054   2.599  0.00934
## InternetServiceFiber optic              1.720069   0.803709   2.140  0.03234
## InternetServiceNo                      -1.622325   0.811846  -1.998  0.04568
## OnlineSecurityNo internet service             NA         NA      NA       NA
## OnlineSecurityYes                      -0.199497   0.179719  -1.110  0.26698
## OnlineBackupNo internet service               NA         NA      NA       NA
## OnlineBackupYes                         0.049975   0.176251   0.284  0.77676
## DeviceProtectionNo internet service           NA         NA      NA       NA
## DeviceProtectionYes                     0.162576   0.177303   0.917  0.35918
## TechSupportNo internet service                NA         NA      NA       NA
## TechSupportYes                         -0.168836   0.181586  -0.930  0.35248
## StreamingTVNo internet service                NA         NA      NA       NA
```

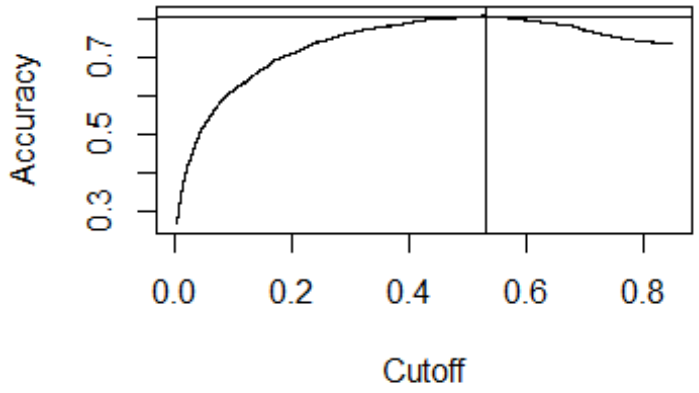
```
## StreamingTVYes                          0.593806   0.328488   1.808  0.07065
## StreamingMoviesNo internet service            NA         NA      NA       NA
## StreamingMoviesYes                      0.608397   0.328840   1.850  0.06429
## ContractOne year                       -0.666321   0.106644  -6.248 4.15e-10
## ContractTwo year                       -1.356836   0.173956  -7.800 6.20e-15
## PaperlessBillingYes                     0.335906   0.074277   4.522 6.12e-06
## PaymentMethodCredit card (automatic)   -0.086598   0.114085  -0.759  0.44782
## PaymentMethodElectronic check           0.314319   0.094582   3.323  0.00089
## PaymentMethodMailed check              -0.005299   0.113719  -0.047  0.96283
## MonthlyCharges                         -0.032716   0.031940  -1.024  0.30570
##
## (Intercept)
## genderMale
## SeniorCitizen                          *
## PartnerYes
## DependentsYes                          .
## tenure                                 ***
## PhoneServiceYes
## MultipleLinesNo phone service
## MultipleLinesYes                       **
## InternetServiceFiber optic             *
## InternetServiceNo                      *
## OnlineSecurityNo internet service
## OnlineSecurityYes
## OnlineBackupNo internet service
## OnlineBackupYes
## DeviceProtectionNo internet service
## DeviceProtectionYes
## TechSupportNo internet service
## TechSupportYes
## StreamingTVNo internet service
## StreamingTVYes                         .
## StreamingMoviesNo internet service
## StreamingMoviesYes                     .
## ContractOne year                       ***
## ContractTwo year                       ***
## PaperlessBillingYes                    ***
## PaymentMethodCredit card (automatic)
## PaymentMethodElectronic check          ***
## PaymentMethodMailed check
## MonthlyCharges
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8150.1  on 7042  degrees of freedom
## Residual deviance: 5851.0  on 7020  degrees of freedom
## AIC: 5897
```

```
##
## Number of Fisher Scoring iterations: 6
```

The most significant variables were SeniorCitizen, tenure, MultipleLines, InternetService, Contract, PaperlessBilling, and PaymentMethod.

ROC

```
##accuracy
plot(performance(ROCRpred, "acc"))
abline(h=0.805, v=0.53)
```

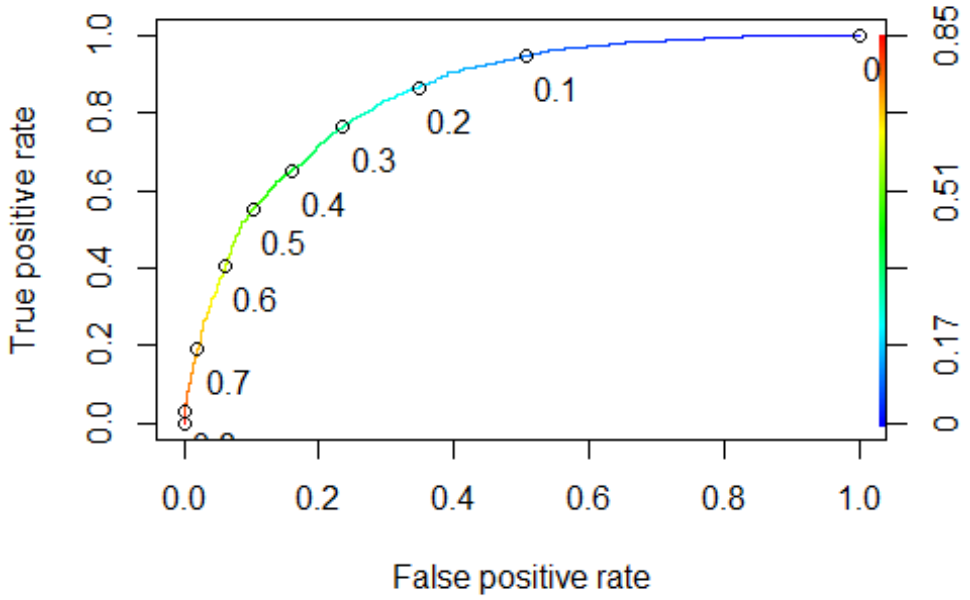



```
table(telecomTrain$Churn,predictTrainn>0.53)

##
##       FALSE TRUE
##   No   3545  335
##   Yes   676  726
```

With a cutoff of 0.53.The true positive rate is 726/(335+726)= 0.6843, So 68.43% of the time the model can predict a customer will churn and they would churn. While False Positive rate is 676/(676+3545)= 0.1602, so 16.02% the model would predict a customer will churn though they stayed. The accuracy of the model is (3545+726)/(3545+335+676+726)= 0.8086. This model has an accuracy of 80.86%

```
#ROC Curve
ROCRpred<-prediction(predictTrainn,telecomTrain$Churn)
ROCRperf<- performance(ROCRpred, "tpr","fpr")
plot(ROCRperf, colorize=TRUE, print.cutoffs.at=seq(0,1,0.1),text.adj=c(-
0.2,1.7))
```

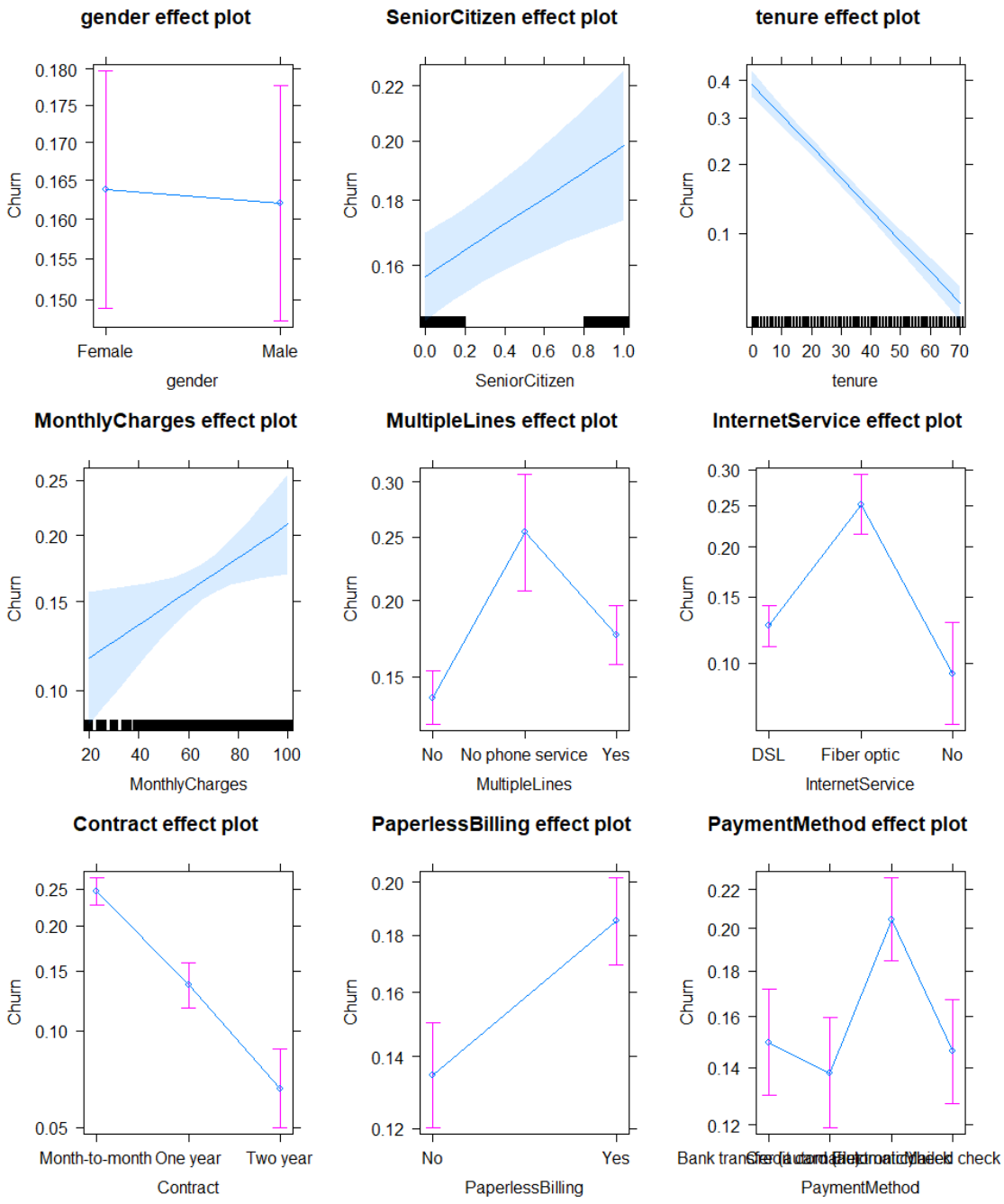



```
AUC<-performance(ROCRpred, "auc")
AUC<-unlist(slot(AUC,"y.values"))
AUC<-round(AUC,4)
AUC

## [1] 0.8475
```

The area under the curve of our model is 0.8475. Our model has an accuracy of 84.75% which is really good.
#Predicting Churning on Multiple Variables

```
churnmodel1 <-
glm(Churn~gender+SeniorCitizen+tenure+MonthlyCharges+MultipleLines+InternetSe
rvice+Contract+PaperlessBilling+PaymentMethod,data=telecom,
family="binomial")
plot(allEffects(churnmodel1))
```



Conclusion

The telecom customer churn analysis depicts various interesting results some of which include:

1. Females are ~0.2% more likely to churn.
2. Senior Citizens are ~4% likely to churn.
3. Customers with tenure of 0 months are ~40% more likely to churn compared to customers with tenure of 72 months. Between 0 to 40 months the customer is likely to

churn. The company should focus on their services during this period.
4. Higher monthly charges to customers are more likely to churn. A customer paying $100 monthly is 1.75x more likely to churn than that of a customer paying ~$20 per month.
5. Customers with multiple lines are 1.3x more likely to churn compared to people with no multiple lines (single line). Customers with no phone service are 1.9x more likely to churn in comparison with single line service.
6. Customers with Fiber Optics are 2.7x more likely to churn in comparison to customers with no internet service. While DSL customers are 1.4x likely to churn compared to customers with no internet service.
7. Month to Month customers are 3.5x more likely to churn than a two year contracted customer. While a one year contracted customer is 1.8x more likely to churn than a two year contracted customer.
8. Customers with paperless billing are 1.37x more likely to churn than those receiving their monthly bill in the mail.
9. Customers paying with Electronic Check are 1.4x more likely to churn in comparison to customers paying in credit card. While customers paying by bank transfer were 1.07x and customers paying by mailed check was 1.03x more likely to churn in comparison to customers paying in credit card.

Recommendations

More research would need to be done to see if these trends are specific to this data set or can be used to speak of other telecom data sets as well. Addition of detailed variables to this data such as price of each service, the location of the customer, demography, and age of customer would help gain further insight.