workshop 4.4 solution

Veerasak Kritsanapraphan

1/22/2020

library(party)  
telco <- read.csv("WA\_Fn-UseC\_-Telco-Customer-Churn.csv", header=TRUE)

str(telco)

## 'data.frame': 7043 obs. of 21 variables:  
## $ customerID : Factor w/ 7043 levels "0002-ORFBO","0003-MKNFE",..: 5376 3963 2565 5536 6512 6552 1003 4771 5605 4535 ...  
## $ gender : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 1 2 1 1 2 ...  
## $ SeniorCitizen : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Partner : Factor w/ 2 levels "No","Yes": 2 1 1 1 1 1 1 1 2 1 ...  
## $ Dependents : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 2 1 1 2 ...  
## $ tenure : int 1 34 2 45 2 8 22 10 28 62 ...  
## $ PhoneService : Factor w/ 2 levels "No","Yes": 1 2 2 1 2 2 2 1 2 2 ...  
## $ MultipleLines : Factor w/ 3 levels "No","No phone service",..: 2 1 1 2 1 3 3 2 3 1 ...  
## $ InternetService : Factor w/ 3 levels "DSL","Fiber optic",..: 1 1 1 1 2 2 2 1 2 1 ...  
## $ OnlineSecurity : Factor w/ 3 levels "No","No internet service",..: 1 3 3 3 1 1 1 3 1 3 ...  
## $ OnlineBackup : Factor w/ 3 levels "No","No internet service",..: 3 1 3 1 1 1 3 1 1 3 ...  
## $ DeviceProtection: Factor w/ 3 levels "No","No internet service",..: 1 3 1 3 1 3 1 1 3 1 ...  
## $ TechSupport : Factor w/ 3 levels "No","No internet service",..: 1 1 1 3 1 1 1 1 3 1 ...  
## $ StreamingTV : Factor w/ 3 levels "No","No internet service",..: 1 1 1 1 1 3 3 1 3 1 ...  
## $ StreamingMovies : Factor w/ 3 levels "No","No internet service",..: 1 1 1 1 1 3 1 1 3 1 ...  
## $ Contract : Factor w/ 3 levels "Month-to-month",..: 1 2 1 2 1 1 1 1 1 2 ...  
## $ PaperlessBilling: Factor w/ 2 levels "No","Yes": 2 1 2 1 2 2 2 1 2 1 ...  
## $ PaymentMethod : Factor w/ 4 levels "Bank transfer (automatic)",..: 3 4 4 1 3 3 2 4 3 1 ...  
## $ MonthlyCharges : num 29.9 57 53.9 42.3 70.7 ...  
## $ TotalCharges : num 29.9 1889.5 108.2 1840.8 151.7 ...  
## $ Churn : Factor w/ 2 levels "No","Yes": 1 1 2 1 2 2 1 1 2 1 ...

## Convert Data

### Data Sampling

index <- sample(2, nrow(telco), replace=TRUE, prob=c(0.7,0.3) )  
traindata <- telco[index==1,]  
testdata <- telco[index==2,]  
sprintf("Number of Record in Training Dataset is %d" , nrow(traindata))

## [1] "Number of Record in Training Dataset is 4960"

sprintf("Number of Record in Testing Dataset is %d" , nrow(testdata))

## [1] "Number of Record in Testing Dataset is 2083"

### Train Model

library(nnet)  
formular <- Churn ~ gender + SeniorCitizen + Partner + tenure + PhoneService + MultipleLines + OnlineBackup + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges + TotalCharges  
multi\_model <- multinom(formular, data = traindata)

## # weights: 19 (18 variable)  
## initial value 3430.385397   
## iter 10 value 2291.791052  
## iter 20 value 2072.672041  
## final value 2064.951747   
## converged

summary(multi\_model)

## Call:  
## multinom(formula = formular, data = traindata)  
##   
## Coefficients:  
## Values Std. Err.  
## (Intercept) -0.8827911006 3.196908e-02  
## genderMale -0.0001499918 7.514223e-02  
## SeniorCitizen 0.4030616205 9.525999e-02  
## PartnerYes -0.0575665293 8.238771e-02  
## tenure -0.0615073321 6.297406e-03  
## PhoneServiceYes -1.0211526824 6.202402e-02  
## MultipleLinesNo phone service 0.1383615818 4.740461e-02  
## MultipleLinesYes 0.1105511000 8.919735e-02  
## OnlineBackupNo internet service 0.1028685138 4.550307e-02  
## OnlineBackupYes -0.3675372629 9.114570e-02  
## ContractOne year -0.8470061738 1.719837e-02  
## ContractTwo year -1.5610077949 4.290967e-03  
## PaperlessBillingYes 0.4367368610 8.296244e-02  
## PaymentMethodCredit card (automatic) -0.1224819849 6.806469e-02  
## PaymentMethodElectronic check 0.3709481782 6.479265e-02  
## PaymentMethodMailed check -0.1305113459 7.058613e-02  
## MonthlyCharges 0.0246010691 1.875810e-03  
## TotalCharges 0.0003027162 6.971678e-05  
##   
## Residual Deviance: 4129.903   
## AIC: 4163.903

### Evaluate Model

confusionmatrix <- table(predict(multi\_model, newdata=traindata), traindata$Churn)  
confusionmatrix

##   
## No Yes  
## No 3293 612  
## Yes 366 678

normal\_train\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])  
print(normal\_train\_accuracy)

## [1] 0.8023843

### Evaluate Testing Model

confusionmatrix <- table(predict(multi\_model, newdata=testdata), testdata$Churn)  
confusionmatrix

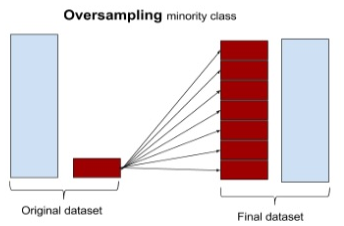
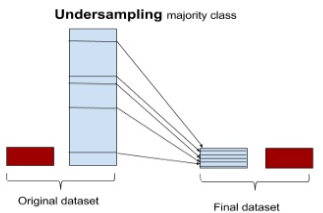
##   
## No Yes  
## No 1345 291  
## Yes 159 288

normal\_test\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])  
print(normal\_test\_accuracy)

## [1] 0.7839654

### Handling Imbalanced Data

* Over Sampling
* Under Sampling
* Both Sampling
* ROSE Sampling
* Synthetic Minority Over-Sampling Technique (SMOTE) Sampling

## Resample Data

library(DMwR)

## Loading required package: lattice

library(ROSE)

## Loaded ROSE 0.0-3

print('Number of transactions in dataset before applying sampling methods')

## [1] "Number of transactions in dataset before applying sampling methods"

print(table(telco$Churn))

##   
## No Yes   
## 5174 1869

# Oversampling,   
# Here N= 5174\*2  
over\_sample\_data <- ovun.sample(Churn ~ ., data = telco, method="over", N=10348)$data  
print('Number of transactions in dataset after applying Over sampling method')

## [1] "Number of transactions in dataset after applying Over sampling method"

print(table(over\_sample\_data$Churn))

##   
## No Yes   
## 5163 5185

# Undersampling, N=1869\*2.   
under\_sample\_data <- ovun.sample(Churn ~ ., data = telco, method="under", N=3738)$data  
print('Number of transactions in dataset after applying Under sampling method')

## [1] "Number of transactions in dataset after applying Under sampling method"

print(table(under\_sample\_data$Churn))

##   
## No Yes   
## 1869 1869

# Mixed Sampling, apply both under sampling and over sampling on this imbalanced data  
both\_sample\_data <- ovun.sample(Churn ~ ., data = telco, method="both", p=0.5, seed=222, N=5174)$data  
print('Number of transactions in dataset after applying Mixed sampling method')

## [1] "Number of transactions in dataset after applying Mixed sampling method"

print(table(both\_sample\_data$Churn))

##   
## No Yes   
## 2614 2560

# ROSE Sampling, this helps us to generate data synthetically. It generates artificial datas instead of dulicate data.  
rose\_sample\_data <- ROSE(Churn ~ ., data = telco, seed=111)$data  
print('Number of transactions in dataset after applying ROSE sampling method')

## [1] "Number of transactions in dataset after applying ROSE sampling method"

print(table(rose\_sample\_data$Churn))

##   
## No Yes   
## 3526 3506

# SMOTE(Synthetic Minority Over-sampling Technique) Sampling  
# formula - relates how our dependent variable acts based on other independent variable.  
# data - input data  
# perc.over - controls the size of Minority class  
# perc.under - controls the size of Majority class  
smote\_sample\_data <- SMOTE(Churn ~ ., data = telco, perc.over = 100, perc.under=200)  
print('Number of transactions in dataset after applying SMOTE sampling method')

## [1] "Number of transactions in dataset after applying SMOTE sampling method"

print(table(smote\_sample\_data$Churn))

##   
## No Yes   
## 3738 3738

### Data Sampling (Over Sample)

index <- sample(2, nrow(over\_sample\_data), replace=TRUE, prob=c(0.7,0.3) )  
traindata <- over\_sample\_data[index==1,]  
testdata <- over\_sample\_data[index==2,]  
sprintf("Number of Record in Training Dataset is %d" , nrow(traindata))

## [1] "Number of Record in Training Dataset is 7205"

sprintf("Number of Record in Testing Dataset is %d" , nrow(testdata))

## [1] "Number of Record in Testing Dataset is 3143"

### Train Model (Over Sample)

library(nnet)  
formular <- Churn ~ gender + SeniorCitizen + Partner + tenure + PhoneService + MultipleLines + OnlineBackup + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges + TotalCharges  
multi\_model <- multinom(formular, data = traindata)

## # weights: 19 (18 variable)  
## initial value 4994.125436   
## iter 10 value 3773.993528  
## iter 20 value 3508.181226  
## final value 3495.556460   
## converged

summary(multi\_model)

## Call:  
## multinom(formula = formular, data = traindata)  
##   
## Coefficients:  
## Values Std. Err.  
## (Intercept) -0.2819628823 2.401052e-02  
## genderMale 0.0357315779 5.746850e-02  
## SeniorCitizen 0.3749455176 7.387791e-02  
## PartnerYes -0.0757403232 6.365878e-02  
## tenure -0.0541821879 4.338432e-03  
## PhoneServiceYes -0.6170182880 4.198538e-02  
## MultipleLinesNo phone service 0.3350554057 3.207024e-02  
## MultipleLinesYes 0.1847480013 6.704334e-02  
## OnlineBackupNo internet service 0.0216141681 3.487253e-02  
## OnlineBackupYes -0.2714602814 6.978575e-02  
## ContractOne year -0.8389007200 2.617182e-02  
## ContractTwo year -1.9286265116 3.912780e-03  
## PaperlessBillingYes 0.3960152683 6.286284e-02  
## PaymentMethodCredit card (automatic) 0.0002707080 5.387020e-02  
## PaymentMethodElectronic check 0.4060124428 5.012467e-02  
## PaymentMethodMailed check -0.0766753589 5.243440e-02  
## MonthlyCharges 0.0239331248 1.422032e-03  
## TotalCharges 0.0002308864 4.851314e-05  
##   
## Residual Deviance: 6991.113   
## AIC: 7025.113

### Evaluate Model (Over Sample)

confusionmatrix <- table(predict(multi\_model, newdata=traindata), traindata$Churn)  
confusionmatrix

##   
## No Yes  
## No 2557 699  
## Yes 1024 2925

over\_train\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Evaluate Testing Model (Over Sample)

confusionmatrix <- table(predict(multi\_model, newdata=testdata), testdata$Churn)  
confusionmatrix

##   
## No Yes  
## No 1150 314  
## Yes 432 1247

over\_test\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Data Sampling (Under Sample)

index <- sample(2, nrow(under\_sample\_data), replace=TRUE, prob=c(0.7,0.3) )  
traindata <- under\_sample\_data[index==1,]  
testdata <- under\_sample\_data[index==2,]  
sprintf("Number of Record in Training Dataset is %d" , nrow(traindata))

## [1] "Number of Record in Training Dataset is 2566"

sprintf("Number of Record in Testing Dataset is %d" , nrow(testdata))

## [1] "Number of Record in Testing Dataset is 1172"

### Train Model (Under Sample)

library(nnet)  
formular <- Churn ~ gender + SeniorCitizen + Partner + tenure + PhoneService + MultipleLines + OnlineBackup + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges + TotalCharges  
multi\_model <- multinom(formular, data = traindata)

## # weights: 19 (18 variable)  
## initial value 1778.615665   
## iter 10 value 1282.566861  
## iter 20 value 1217.745643  
## final value 1213.163358   
## converged

summary(multi\_model)

## Call:  
## multinom(formula = formular, data = traindata)  
##   
## Coefficients:  
## Values Std. Err.  
## (Intercept) -0.0752857923 0.0401397829  
## genderMale 0.0764177034 0.0969097380  
## SeniorCitizen 0.1640044587 0.1195406855  
## PartnerYes 0.0045318959 0.1083046233  
## tenure -0.0717363581 0.0076403353  
## PhoneServiceYes -0.4349190105 0.0714828224  
## MultipleLinesNo phone service 0.3596332182 0.0523138769  
## MultipleLinesYes 0.1742366489 0.1183299683  
## OnlineBackupNo internet service -0.2204794989 0.0602252295  
## OnlineBackupYes -0.2634589769 0.1211441424  
## ContractOne year -0.9402792876 0.0484952939  
## ContractTwo year -2.1228923904 0.0091780822  
## PaperlessBillingYes 0.2572946990 0.1069181516  
## PaymentMethodCredit card (automatic) -0.0010819831 0.0911156883  
## PaymentMethodElectronic check 0.5698730397 0.0838737250  
## PaymentMethodMailed check 0.0867934488 0.0879203233  
## MonthlyCharges 0.0192441886 0.0024221439  
## TotalCharges 0.0004543532 0.0000845646  
##   
## Residual Deviance: 2426.327   
## AIC: 2460.327

### Evaluate Model (Under Sample)

confusionmatrix <- table(predict(multi\_model, newdata=traindata), traindata$Churn)  
confusionmatrix

##   
## No Yes  
## No 922 222  
## Yes 357 1065

under\_train\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Evaluate Testing Model (Under Sample)

confusionmatrix <- table(predict(multi\_model, newdata=testdata), testdata$Churn)  
confusionmatrix

##   
## No Yes  
## No 412 117  
## Yes 178 465

under\_test\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Data Sampling (Both Sample)

index <- sample(2, nrow(both\_sample\_data), replace=TRUE, prob=c(0.7,0.3) )  
traindata <- both\_sample\_data[index==1,]  
testdata <- both\_sample\_data[index==2,]  
sprintf("Number of Record in Training Dataset is %d" , nrow(traindata))

## [1] "Number of Record in Training Dataset is 3579"

sprintf("Number of Record in Testing Dataset is %d" , nrow(testdata))

## [1] "Number of Record in Testing Dataset is 1595"

### Train Model (Both Sample)

library(nnet)  
formular <- Churn ~ gender + SeniorCitizen + Partner + tenure + PhoneService + MultipleLines + OnlineBackup + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges + TotalCharges  
multi\_model <- multinom(formular, data = traindata)

## # weights: 19 (18 variable)  
## initial value 2480.773759   
## iter 10 value 1834.383259  
## iter 20 value 1782.901684  
## final value 1780.679992   
## converged

summary(multi\_model)

## Call:  
## multinom(formula = formular, data = traindata)  
##   
## Coefficients:  
## Values Std. Err.  
## (Intercept) -0.2281652870 0.0296818695  
## genderMale -0.0254760204 0.0799713289  
## SeniorCitizen 0.2828045515 0.1055036386  
## PartnerYes -0.2275553667 0.0876488307  
## tenure -0.0455999518 0.0056536969  
## PhoneServiceYes -0.4788896781 0.0552980814  
## MultipleLinesNo phone service 0.2507243911 0.0420743570  
## MultipleLinesYes 0.2694728301 0.0935785663  
## OnlineBackupNo internet service -0.0887369033 0.0455686275  
## OnlineBackupYes -0.1729323943 0.0981480893  
## ContractOne year -0.7023083563 0.0620215511  
## ContractTwo year -1.7539053189 0.0192255209  
## PaperlessBillingYes 0.4697610969 0.0873210825  
## PaymentMethodCredit card (automatic) -0.1913244046 0.0693369951  
## PaymentMethodElectronic check 0.3440371009 0.0692280426  
## PaymentMethodMailed check 0.0372471271 0.0711545136  
## MonthlyCharges 0.0212138040 0.0019767554  
## TotalCharges 0.0001412911 0.0000649945  
##   
## Residual Deviance: 3561.36   
## AIC: 3595.36

### Evaluate Model (Both Sample)

confusionmatrix <- table(predict(multi\_model, newdata=traindata), traindata$Churn)  
confusionmatrix

##   
## No Yes  
## No 1242 356  
## Yes 536 1445

both\_train\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Evaluate Testing Model (Both Sample)

confusionmatrix <- table(predict(multi\_model, newdata=testdata), testdata$Churn)  
confusionmatrix

##   
## No Yes  
## No 617 150  
## Yes 219 609

both\_test\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Data Sampling (ROSE Sample)

index <- sample(2, nrow(rose\_sample\_data), replace=TRUE, prob=c(0.7,0.3) )  
traindata <- rose\_sample\_data[index==1,]  
testdata <- rose\_sample\_data[index==2,]  
sprintf("Number of Record in Training Dataset is %d" , nrow(traindata))

## [1] "Number of Record in Training Dataset is 4931"

sprintf("Number of Record in Testing Dataset is %d" , nrow(testdata))

## [1] "Number of Record in Testing Dataset is 2101"

### Train Model (ROSE Sample)

library(nnet)  
formular <- Churn ~ gender + SeniorCitizen + Partner + tenure + PhoneService + MultipleLines + OnlineBackup + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges + TotalCharges  
multi\_model <- multinom(formular, data = traindata)

## # weights: 19 (18 variable)  
## initial value 3417.908747   
## iter 10 value 2540.575265  
## iter 20 value 2459.432098  
## final value 2457.399600   
## converged

summary(multi\_model)

## Call:  
## multinom(formula = formular, data = traindata)  
##   
## Coefficients:  
## Values Std. Err.  
## (Intercept) -1.476069e-01 3.070333e-02  
## genderMale -1.087815e-02 6.683424e-02  
## SeniorCitizen 1.892304e-01 8.662174e-02  
## PartnerYes -3.381506e-01 7.477586e-02  
## tenure -2.768974e-02 2.850130e-03  
## PhoneServiceYes -2.480733e-01 4.694678e-02  
## MultipleLinesNo phone service 1.004664e-01 3.608175e-02  
## MultipleLinesYes 2.346509e-01 7.377484e-02  
## OnlineBackupNo internet service -4.933483e-01 3.574227e-02  
## OnlineBackupYes -2.905570e-01 8.164214e-02  
## ContractOne year -8.637906e-01 3.780081e-02  
## ContractTwo year -1.662808e+00 1.127871e-02  
## PaperlessBillingYes 4.629719e-01 7.158557e-02  
## PaymentMethodCredit card (automatic) -6.758385e-02 6.472494e-02  
## PaymentMethodElectronic check 4.699482e-01 5.945840e-02  
## PaymentMethodMailed check -9.076305e-02 6.131079e-02  
## MonthlyCharges 1.731917e-02 1.541647e-03  
## TotalCharges -6.796091e-06 3.164745e-05  
##   
## Residual Deviance: 4914.799   
## AIC: 4948.799

### Evaluate Model (ROSE Sample)

confusionmatrix <- table(predict(multi\_model, newdata=traindata), traindata$Churn)  
confusionmatrix

##   
## No Yes  
## No 1773 490  
## Yes 700 1968

rose\_train\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Evaluate Testing Model (ROSE Sample)

confusionmatrix <- table(predict(multi\_model, newdata=testdata), testdata$Churn)  
confusionmatrix

##   
## No Yes  
## No 760 221  
## Yes 293 827

rose\_test\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Data Sampling (SMOTE Sample)

index <- sample(2, nrow(smote\_sample\_data), replace=TRUE, prob=c(0.7,0.3) )  
traindata <- smote\_sample\_data[index==1,]  
testdata <- smote\_sample\_data[index==2,]  
sprintf("Number of Record in Training Dataset is %d" , nrow(traindata))

## [1] "Number of Record in Training Dataset is 5176"

sprintf("Number of Record in Testing Dataset is %d" , nrow(testdata))

## [1] "Number of Record in Testing Dataset is 2300"

### Train Model (SMOTE Sample)

library(nnet)  
formular <- Churn ~ gender + SeniorCitizen + Partner + tenure + PhoneService + MultipleLines + OnlineBackup + Contract + PaperlessBilling + PaymentMethod + MonthlyCharges + TotalCharges  
multi\_model <- multinom(formular, data = traindata)

## # weights: 19 (18 variable)  
## initial value 3583.570923   
## iter 10 value 3086.218061  
## iter 20 value 2906.324819  
## final value 2898.820447   
## converged

summary(multi\_model)

## Call:  
## multinom(formula = formular, data = traindata)  
##   
## Coefficients:  
## Values Std. Err.  
## (Intercept) 1.378798e-01 0.025481458  
## genderMale 1.935428e-03 0.062317709  
## SeniorCitizen 2.882915e-01 0.087773036  
## PartnerYes 2.178989e-01 0.065874731  
## tenure -3.805450e-02 0.003670095  
## PhoneServiceYes -3.781425e-01 0.038182368  
## MultipleLinesNo phone service 5.360780e-01 0.028019319  
## MultipleLinesYes 2.920644e-01 0.070917541  
## OnlineBackupNo internet service 3.406061e-01 0.041794031  
## OnlineBackupYes -2.117700e-01 0.077749490  
## ContractOne year -4.119073e-01 0.069403339  
## ContractTwo year -9.700481e-01 0.033482112  
## PaperlessBillingYes 2.384438e-01 0.065477026  
## PaymentMethodCredit card (automatic) -6.255211e-02 0.057706730  
## PaymentMethodElectronic check 3.147062e-01 0.054499874  
## PaymentMethodMailed check -3.727630e-01 0.054544924  
## MonthlyCharges 1.562824e-02 0.001588573  
## TotalCharges -1.483726e-05 0.000047015  
##   
## Residual Deviance: 5797.641   
## AIC: 5833.641

### Evaluate Model (SMOTE Sample)

confusionmatrix <- table(predict(multi\_model, newdata=traindata), traindata$Churn)  
confusionmatrix

##   
## No Yes  
## No 1748 656  
## Yes 856 1910

smote\_train\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Evaluate Testing Model (SMOTE Sample)

confusionmatrix <- table(predict(multi\_model, newdata=testdata), testdata$Churn)  
confusionmatrix

##   
## No Yes  
## No 765 317  
## Yes 362 855

smote\_test\_accuracy <- (confusionmatrix[1,1]+confusionmatrix[2,2])/  
 (confusionmatrix[1,1]+confusionmatrix[2,2]+confusionmatrix[1,2]+confusionmatrix[2,1])

### Compare Accuracy Data

compareconfusion <- data.frame(  
 "Data"= c("Imbalanced", "Over-sample", "Under-sample", "Both-sample", "ROSE sample", "SMOTE sample"),  
 "Train Accuracy"= c(normal\_train\_accuracy, over\_train\_accuracy, under\_train\_accuracy, both\_train\_accuracy, rose\_train\_accuracy, smote\_train\_accuracy),  
 "Test Accuracy"= c(normal\_test\_accuracy, over\_test\_accuracy, under\_test\_accuracy, both\_test\_accuracy, rose\_test\_accuracy, smote\_test\_accuracy)  
)  
  
print(compareconfusion)

## Data Train.Accuracy Test.Accuracy  
## 1 Imbalanced 0.8023843 0.7839654  
## 2 Over-sample 0.7608605 0.7626472  
## 3 Under-sample 0.7743570 0.7482935  
## 4 Both-sample 0.7507684 0.7686520  
## 5 ROSE sample 0.7586696 0.7553546  
## 6 SMOTE sample 0.7075435 0.7046542