|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| > set.seed(1)  > index<-sample(1:nrow(data),0.8\*nrow(data))  > train\_data<-data[index,-21]  > str(train\_data)  'data.frame': 2666 obs. of 20 variables:  $ Account.Length: int 73 64 88 101 3 108 95 133 33 66 ...  $ VMail.Message : int 0 0 45 0 36 0 0 0 0 36 ...  $ Day.Mins : num 254.8 168 80.3 118.6 118.1 ...  $ Eve.Mins : num 143 192 153 200 222 ...  $ Night.Mins : num 153.9 166.5 309.2 53.3 103.9 ...  $ Intl.Mins : num 15 10.1 12.8 11.5 11.9 13.8 12.8 3.1 7.5 12.1 ...  $ CustServ.Calls: int 2 2 2 1 2 2 2 2 2 1 ...  $ Churn : int 0 0 0 0 0 1 0 0 0 0 ...  $ Int.l.Plan : int 0 0 0 0 0 1 0 0 0 0 ...  $ VMail.Plan : int 0 0 1 0 1 0 0 0 0 1 ...  $ Day.Calls : int 85 116 140 89 117 114 95 67 65 76 ...  $ Day.Charge : num 43.3 28.6 13.6 20.2 20.1 ...  $ Eve.Calls : int 80 94 101 97 125 70 101 133 96 77 ...  $ Eve.Charge : num 12.2 16.4 13 17 18.8 ...  $ Night.Calls : int 102 98 123 61 89 85 113 95 82 117 ...  $ Night.Charge : num 6.93 7.49 13.91 2.4 4.68 ...  $ Intl.Calls : int 7 3 3 5 6 3 4 1 2 2 ...  $ Intl.Charge : num 4.05 2.73 3.46 3.11 3.21 3.73 3.46 0.84 2.03 3.27 ...  $ State : Factor w/ 51 levels "AK","AL","AR",..: 9 44 20 38 7 23 9 19 7 25 ...  $ Area.Code : int 415 415 408 415 415 408 510 415 510 415 ...  > test\_data<-data[-index,c(-8,-21)]  > str(test\_data)  'data.frame': 667 obs. of 19 variables:  $ Account.Length: int 137 121 65 76 57 57 78 59 10 96 ...  $ VMail.Message : int 0 24 0 33 39 25 0 28 0 0 ...  $ Day.Mins : num 243 218 129 190 213 ...  $ Eve.Mins : num 121 348 228 213 191 ...  $ Night.Mins : num 163 213 209 166 183 ...  $ Intl.Mins : num 12.2 7.5 12.7 10 9.5 8.3 10 8.5 11.4 9.3 ...  $ CustServ.Calls: int 0 3 4 1 0 0 1 2 2 2 ...  $ Int.l.Plan : int 0 0 0 0 0 0 0 0 0 0 ...  $ VMail.Plan : int 0 1 0 1 1 1 0 1 0 0 ...  $ Day.Calls : int 114 88 137 66 115 94 64 97 112 117 ...  $ Day.Charge : num 41.4 37.1 21.9 32.2 36.2 ...  $ Eve.Calls : int 110 108 83 65 112 75 116 92 66 67 ...  $ Eve.Charge : num 10.3 29.6 19.4 18.1 16.2 ...  $ Night.Calls : int 104 118 111 108 115 116 108 116 57 68 ...  $ Night.Charge : num 7.32 9.57 9.4 7.46 8.22 ...  $ Intl.Calls : int 5 7 6 5 3 4 5 5 6 5 ...  $ Intl.Charge : num 3.29 2.03 3.43 2.7 2.57 2.24 2.7 2.3 3.08 2.51 ...  $ State : Factor w/ 51 levels "AK","AL","AR",..: 32 20 16 46 51 37 20 38 46 45 ...  $ Area.Code : int 415 510 415 510 408 408 415 408 408 415 ...  > glm\_model<-glm(Churn~.,family=binomial,data=train\_data)  > summary(glm\_model)  Call:  glm(formula = Churn ~ ., family = binomial, data = train\_data)  Deviance Residuals:  Min 1Q Median 3Q Max  -2.1036 -0.4969 -0.3060 -0.1539 3.0298  Coefficients:  Estimate Std. Error z value Pr(>|z|)  (Intercept) -1.019e+01 1.338e+00 -7.614 2.66e-14 \*\*\*  Account.Length 2.261e-03 1.637e-03 1.381 0.16729  VMail.Message 3.641e-02 2.137e-02 1.704 0.08836 .  Day.Mins 2.168e+00 3.825e+00 0.567 0.57090  Eve.Mins -1.824e-01 1.941e+00 -0.094 0.92516  Night.Mins 2.337e-01 1.017e+00 0.230 0.81820  Intl.Mins -1.184e+01 6.224e+00 -1.902 0.05722 .  CustServ.Calls 5.040e-01 4.554e-02 11.068 < 2e-16 \*\*\*  Int.l.Plan 2.211e+00 1.742e-01 12.692 < 2e-16 \*\*\*  VMail.Plan -2.215e+00 6.820e-01 -3.247 0.00117 \*\*  Day.Calls 2.899e-03 3.281e-03 0.884 0.37696  Day.Charge -1.267e+01 2.250e+01 -0.563 0.57342  Eve.Calls -4.082e-04 3.254e-03 -0.125 0.90019  Eve.Charge 2.232e+00 2.284e+01 0.098 0.92215  Night.Calls 1.446e-03 3.325e-03 0.435 0.66368  Night.Charge -5.105e+00 2.259e+01 -0.226 0.82123  Intl.Calls -8.938e-02 2.893e-02 -3.089 0.00201 \*\*  Intl.Charge 4.414e+01 2.305e+01 1.915 0.05549 .  StateAL 6.894e-01 8.980e-01 0.768 0.44267  StateAR 1.311e+00 8.815e-01 1.487 0.13709  StateAZ 5.685e-01 9.618e-01 0.591 0.55450  StateCA 2.093e+00 9.342e-01 2.241 0.02504 \*  StateCO 9.463e-01 9.170e-01 1.032 0.30209  StateCT 1.281e+00 8.667e-01 1.478 0.13953  StateDC 1.252e+00 9.290e-01 1.348 0.17779  StateDE 9.292e-01 9.117e-01 1.019 0.30809  StateFL 8.914e-01 8.985e-01 0.992 0.32117  StateGA 1.064e+00 9.357e-01 1.137 0.25564  StateHI 3.303e-01 1.009e+00 0.328 0.74327  StateIA 7.649e-01 1.014e+00 0.754 0.45082  StateID 1.318e+00 8.878e-01 1.485 0.13752  StateIL -9.365e-03 9.794e-01 -0.010 0.99237  StateIN 4.707e-01 9.345e-01 0.504 0.61446  StateKS 1.612e+00 8.635e-01 1.866 0.06200 .  StateKY 1.129e+00 9.212e-01 1.225 0.22052  StateLA 1.228e+00 9.526e-01 1.290 0.19720  StateMA 1.563e+00 8.831e-01 1.769 0.07682 .  StateMD 1.603e+00 8.608e-01 1.862 0.06254 .  StateME 1.714e+00 8.751e-01 1.959 0.05011 .  StateMI 1.801e+00 8.658e-01 2.080 0.03753 \*  StateMN 1.431e+00 8.651e-01 1.654 0.09809 .  StateMO 5.617e-01 9.621e-01 0.584 0.55932  StateMS 2.116e+00 8.678e-01 2.438 0.01475 \*  StateMT 2.082e+00 8.605e-01 2.419 0.01556 \*  StateNC 4.477e-01 9.269e-01 0.483 0.62905  StateND 7.776e-01 9.388e-01 0.828 0.40753  StateNE 3.269e-01 1.003e+00 0.326 0.74439  StateNH 1.490e+00 9.208e-01 1.619 0.10553  StateNJ 1.941e+00 8.487e-01 2.287 0.02222 \*  StateNM 5.213e-01 9.606e-01 0.543 0.58739  StateNV 1.823e+00 8.697e-01 2.097 0.03603 \*  StateNY 1.526e+00 8.568e-01 1.782 0.07482 .  StateOH 1.042e+00 8.868e-01 1.175 0.24000  StateOK 9.720e-01 9.060e-01 1.073 0.28330  StateOR 8.553e-01 9.017e-01 0.949 0.34286  StatePA 1.399e+00 9.424e-01 1.484 0.13768  StateRI -8.922e-01 1.135e+00 -0.786 0.43186  StateSC 1.946e+00 8.840e-01 2.202 0.02770 \*  StateSD 9.870e-01 9.124e-01 1.082 0.27935  StateTN 2.125e-01 1.017e+00 0.209 0.83446  StateTX 1.986e+00 8.453e-01 2.349 0.01882 \*  StateUT 1.160e+00 9.075e-01 1.279 0.20103  StateVA 7.655e-02 9.696e-01 0.079 0.93708  StateVT 1.631e-01 9.354e-01 0.174 0.86161  StateWA 1.404e+00 8.815e-01 1.593 0.11110  StateWI 8.627e-01 9.054e-01 0.953 0.34069  StateWV 9.947e-01 8.709e-01 1.142 0.25342  StateWY 8.576e-01 8.852e-01 0.969 0.33266  Area.Code 8.878e-05 1.528e-03 0.058 0.95367  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for binomial family taken to be 1)  Null deviance: 2198.0 on 2665 degrees of freedom  Residual deviance: 1632.2 on 2597 degrees of freedom  AIC: 1770.2  Number of Fisher Scoring iterations: 6  > exp(glm\_model$coefficients)  (Intercept) Account.Length VMail.Message Day.Mins Eve.Mins Night.Mins Intl.Mins  3.773130e-05 1.002263e+00 1.037085e+00 8.739125e+00 8.332902e-01 1.263247e+00 7.238132e-06  CustServ.Calls Int.l.Plan VMail.Plan Day.Calls Day.Charge Eve.Calls Eve.Charge  1.655335e+00 9.128486e+00 1.092075e-01 1.002903e+00 3.148077e-06 9.995919e-01 9.320343e+00  Night.Calls Night.Charge Intl.Calls Intl.Charge StateAL StateAR StateAZ  1.001447e+00 6.067665e-03 9.144949e-01 1.481880e+19 1.992459e+00 3.708028e+00 1.765529e+00  StateCA StateCO StateCT StateDC StateDE StateFL StateGA  8.112565e+00 2.576264e+00 3.598823e+00 3.497134e+00 2.532603e+00 2.438510e+00 2.897042e+00  StateHI StateIA StateID StateIL StateIN StateKS StateKY  1.391430e+00 2.148801e+00 3.737693e+00 9.906788e-01 1.601157e+00 5.010718e+00 3.091431e+00  StateLA StateMA StateMD StateME StateMI StateMN StateMO  3.415847e+00 4.771397e+00 4.968510e+00 5.553133e+00 6.054529e+00 4.183253e+00 1.753711e+00  StateMS StateMT StateNC StateND StateNE StateNH StateNJ  8.296937e+00 8.016832e+00 1.564775e+00 2.176213e+00 1.386625e+00 4.438948e+00 6.963518e+00  StateNM StateNV StateNY StateOH StateOK StateOR StatePA  1.684146e+00 6.192454e+00 4.601859e+00 2.834896e+00 2.643277e+00 2.352160e+00 4.050920e+00  StateRI StateSC StateSD StateTN StateTX StateUT StateVA  4.097484e-01 7.001605e+00 2.683216e+00 1.236731e+00 7.283897e+00 3.190946e+00 1.079551e+00  StateVT StateWA StateWI StateWV StateWY Area.Code  1.177111e+00 4.073229e+00 2.369549e+00 2.703904e+00 2.357454e+00 1.000089e+00  > pred <- predict(glm\_model,test\_data,type="response")  > outcome=floor(pred+0.5)  > table(outcome)  outcome  0 1  634 33  > ttt=table(data$Churn[-index] , outcome )  > ttt  outcome  0 1  0 557 11  1 77 22  # model2 removing state#  > glm\_model1<-glm(Churn~. -State ,family=binomial,data=train\_data)  > glm\_model1  Call: glm(formula = Churn ~ . - State, family = binomial, data = train\_data)  Coefficients:  (Intercept) Account.Length VMail.Message Day.Mins Eve.Mins Night.Mins Intl.Mins CustServ.Calls  -8.712e+00 1.621e-03 3.168e-02 2.959e+00 -3.866e-01 1.859e-01 -1.072e+01 4.776e-01  Int.l.Plan VMail.Plan Day.Calls Day.Charge Eve.Calls Eve.Charge Night.Calls Night.Charge  2.032e+00 -1.969e+00 2.659e-03 -1.732e+01 4.356e-05 4.627e+00 1.884e-03 -4.047e+00  Intl.Calls Intl.Charge Area.Code  -9.125e-02 4.003e+01 4.128e-05  Degrees of Freedom: 2665 Total (i.e. Null); 2647 Residual  Null Deviance: 2198  Residual Deviance: 1717 AIC: 1755  > summary(glm\_model1)  Call:  glm(formula = Churn ~ . - State, family = binomial, data = train\_data)  Deviance Residuals:  Min 1Q Median 3Q Max  -2.0592 -0.5131 -0.3365 -0.1900 3.1713  Coefficients:  Estimate Std. Error z value Pr(>|z|)  (Intercept) -8.712e+00 1.041e+00 -8.367 < 2e-16 \*\*\*  Account.Length 1.621e-03 1.573e-03 1.030 0.30278  VMail.Message 3.168e-02 2.056e-02 1.541 0.12342  Day.Mins 2.959e+00 3.666e+00 0.807 0.41961  Eve.Mins -3.866e-01 1.849e+00 -0.209 0.83441  Night.Mins 1.859e-01 9.799e-01 0.190 0.84955  Intl.Mins -1.072e+01 5.954e+00 -1.801 0.07177 .  CustServ.Calls 4.776e-01 4.343e-02 10.996 < 2e-16 \*\*\*  Int.l.Plan 2.032e+00 1.637e-01 12.411 < 2e-16 \*\*\*  VMail.Plan -1.969e+00 6.552e-01 -3.006 0.00265 \*\*  Day.Calls 2.659e-03 3.130e-03 0.850 0.39553  Day.Charge -1.732e+01 2.157e+01 -0.803 0.42177  Eve.Calls 4.356e-05 3.108e-03 0.014 0.98882  Eve.Charge 4.627e+00 2.176e+01 0.213 0.83159  Night.Calls 1.884e-03 3.211e-03 0.587 0.55742  Night.Charge -4.047e+00 2.177e+01 -0.186 0.85254  Intl.Calls -9.125e-02 2.803e-02 -3.255 0.00113 \*\*  Intl.Charge 4.003e+01 2.205e+01 1.815 0.06949 .  Area.Code 4.128e-05 1.474e-03 0.028 0.97766  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for binomial family taken to be 1)  Null deviance: 2198.0 on 2665 degrees of freedom  Residual deviance: 1717.3 on 2647 degrees of freedom  AIC: 1755.3  Number of Fisher Scoring iterations: 6  > pred1 <- predict(glm\_model1,test\_data,type="response")  > outcome1 <- floor(pred1+0.5)  > table(outcome1)  outcome1  0 1  636 31  > ttt1<- table(data$Churn[-index],outcome1)  > ttt1  outcome1  0 1  0 556 12  1 80 19  > data\_1<-data[data$Churn==1,]  > ind\_1<-sample(rownames(data\_1),483)  > data\_0<-data[data$Churn==0,]  > ind\_0<-sample(rownames(data\_0),483)  > train\_data1<-data[c(ind\_1,ind\_0), -21 ]  > table(train\_data1$Churn)  0 1  483 483  ####  > glm\_model2<-glm(Churn~. -State ,family=binomial,data=train\_data1)  > glm\_model2  Call: glm(formula = Churn ~ . - State, family = binomial, data = train\_data1)  Coefficients:  (Intercept) Account.Length VMail.Message Day.Mins Eve.Mins Night.Mins Intl.Mins CustServ.Calls  -6.796e+00 1.210e-03 2.396e-03 -6.467e-01 2.133e+00 2.753e-01 -4.983e+00 6.191e-01  Int.l.Plan VMail.Plan Day.Calls Day.Charge Eve.Calls Eve.Charge Night.Calls Night.Charge  2.107e+00 -8.959e-01 3.333e-03 3.881e+00 3.304e-03 -2.502e+01 -3.265e-03 -6.074e+00  Intl.Calls Intl.Charge Area.Code  -8.082e-02 1.874e+01 8.147e-04  Degrees of Freedom: 965 Total (i.e. Null); 947 Residual  Null Deviance: 1339  Residual Deviance: 1009 AIC: 1047  > summary(glm\_model2)  Call:  glm(formula = Churn ~ . - State, family = binomial, data = train\_data1)  Deviance Residuals:  Min 1Q Median 3Q Max  -2.72207 -0.82007 -0.04749 0.86912 2.58151  Coefficients:  Estimate Std. Error z value Pr(>|z|)  (Intercept) -6.796e+00 1.248e+00 -5.445 5.19e-08 \*\*\*  Account.Length 1.210e-03 1.992e-03 0.607 0.5437  VMail.Message 2.396e-03 2.405e-02 0.100 0.9207  Day.Mins -6.467e-01 4.637e+00 -0.139 0.8891  Eve.Mins 2.133e+00 2.267e+00 0.941 0.3468  Night.Mins 2.753e-01 1.215e+00 0.227 0.8208  Intl.Mins -4.983e+00 7.371e+00 -0.676 0.4991  CustServ.Calls 6.191e-01 5.793e-02 10.687 < 2e-16 \*\*\*  Int.l.Plan 2.107e+00 2.342e-01 8.997 < 2e-16 \*\*\*  VMail.Plan -8.959e-01 7.688e-01 -1.165 0.2439  Day.Calls 3.333e-03 3.798e-03 0.878 0.3802  Day.Charge 3.881e+00 2.728e+01 0.142 0.8869  Eve.Calls 3.304e-03 3.812e-03 0.867 0.3861  Eve.Charge -2.502e+01 2.667e+01 -0.938 0.3482  Night.Calls -3.265e-03 3.972e-03 -0.822 0.4111  Night.Charge -6.074e+00 2.700e+01 -0.225 0.8220  Intl.Calls -8.082e-02 3.272e-02 -2.470 0.0135 \*  Intl.Charge 1.874e+01 2.730e+01 0.686 0.4924  Area.Code 8.147e-04 1.842e-03 0.442 0.6583  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for binomial family taken to be 1)  Null deviance: 1339.2 on 965 degrees of freedom  Residual deviance: 1008.5 on 947 degrees of freedom  AIC: 1046.5  Number of Fisher Scoring iterations: 4  > pred2 <- predict(glm\_model2,test\_data,type="response")  > outcome2 <- floor(pred2+0.5)  > table(outcome2)  outcome2  0 1  467 200  > ttt2<- table(data$Churn[-index],outcome2)  > ttt2  outcome2  0 1  0 442 126  1 25 74  > x1 <- step(glm\_model2)  Start: AIC=1046.5  Churn ~ (Account.Length + VMail.Message + Day.Mins + Eve.Mins +  Night.Mins + Intl.Mins + CustServ.Calls + Int.l.Plan + VMail.Plan +  Day.Calls + Day.Charge + Eve.Calls + Eve.Charge + Night.Calls +  Night.Charge + Intl.Calls + Intl.Charge + State + Area.Code) -  State  Df Deviance AIC  - VMail.Message 1 1008.5 1044.5  - Day.Mins 1 1008.5 1044.5  - Day.Charge 1 1008.5 1044.5  - Night.Charge 1 1008.6 1044.6  - Night.Mins 1 1008.6 1044.6  - Area.Code 1 1008.7 1044.7  - Account.Length 1 1008.9 1044.9  - Intl.Mins 1 1009.0 1045.0  - Intl.Charge 1 1009.0 1045.0  - Night.Calls 1 1009.2 1045.2  - Eve.Calls 1 1009.3 1045.3  - Day.Calls 1 1009.3 1045.3  - Eve.Charge 1 1009.4 1045.4  - Eve.Mins 1 1009.4 1045.4  - VMail.Plan 1 1009.9 1045.9  <none> 1008.5 1046.5  - Intl.Calls 1 1014.7 1050.7  - Int.l.Plan 1 1107.8 1143.8  - CustServ.Calls 1 1150.7 1186.7  Step: AIC=1044.51  Churn ~ Account.Length + Day.Mins + Eve.Mins + Night.Mins + Intl.Mins +  CustServ.Calls + Int.l.Plan + VMail.Plan + Day.Calls + Day.Charge +  Eve.Calls + Eve.Charge + Night.Calls + Night.Charge + Intl.Calls +  Intl.Charge + Area.Code  Df Deviance AIC  - Day.Mins 1 1008.5 1042.5  - Day.Charge 1 1008.5 1042.5  - Night.Charge 1 1008.6 1042.6  - Night.Mins 1 1008.6 1042.6  - Area.Code 1 1008.7 1042.7  - Account.Length 1 1008.9 1042.9  - Intl.Mins 1 1009.0 1043.0  - Intl.Charge 1 1009.0 1043.0  - Night.Calls 1 1009.2 1043.2  - Eve.Calls 1 1009.3 1043.3  - Day.Calls 1 1009.3 1043.3  - Eve.Charge 1 1009.4 1043.4  - Eve.Mins 1 1009.4 1043.4  <none> 1008.5 1044.5  - Intl.Calls 1 1014.7 1048.7  - VMail.Plan 1 1027.7 1061.7  - Int.l.Plan 1 1107.8 1141.8  - CustServ.Calls 1 1150.7 1184.7  Step: AIC=1042.53  Churn ~ Account.Length + Eve.Mins + Night.Mins + Intl.Mins +  CustServ.Calls + Int.l.Plan + VMail.Plan + Day.Calls + Day.Charge +  Eve.Calls + Eve.Charge + Night.Calls + Night.Charge + Intl.Calls +  Intl.Charge + Area.Code  Df Deviance AIC  - Night.Charge 1 1008.6 1040.6  - Night.Mins 1 1008.6 1040.6  - Area.Code 1 1008.7 1040.7  - Account.Length 1 1008.9 1040.9  - Intl.Mins 1 1009.0 1041.0  - Intl.Charge 1 1009.0 1041.0  - Night.Calls 1 1009.2 1041.2  - Eve.Calls 1 1009.3 1041.3  - Day.Calls 1 1009.3 1041.3  - Eve.Charge 1 1009.4 1041.4  - Eve.Mins 1 1009.4 1041.4  <none> 1008.5 1042.5  - Intl.Calls 1 1014.7 1046.7  - VMail.Plan 1 1027.7 1059.7  - Int.l.Plan 1 1107.9 1139.9  - Day.Charge 1 1108.0 1140.0  - CustServ.Calls 1 1150.9 1182.9  Step: AIC=1040.59  Churn ~ Account.Length + Eve.Mins + Night.Mins + Intl.Mins +  CustServ.Calls + Int.l.Plan + VMail.Plan + Day.Calls + Day.Charge +  Eve.Calls + Eve.Charge + Night.Calls + Intl.Calls + Intl.Charge +  Area.Code  Df Deviance AIC  - Area.Code 1 1008.8 1038.8  - Account.Length 1 1009.0 1039.0  - Intl.Mins 1 1009.0 1039.0  - Intl.Charge 1 1009.1 1039.1  - Night.Calls 1 1009.3 1039.3  - Eve.Calls 1 1009.3 1039.3  - Day.Calls 1 1009.4 1039.4  - Eve.Charge 1 1009.5 1039.5  - Eve.Mins 1 1009.5 1039.5  - Night.Mins 1 1010.0 1040.0  <none> 1008.6 1040.6  - Intl.Calls 1 1014.8 1044.8  - VMail.Plan 1 1027.7 1057.7  - Int.l.Plan 1 1108.0 1138.0  - Day.Charge 1 1108.0 1138.0  - CustServ.Calls 1 1150.9 1180.9  Step: AIC=1038.8  Churn ~ Account.Length + Eve.Mins + Night.Mins + Intl.Mins +  CustServ.Calls + Int.l.Plan + VMail.Plan + Day.Calls + Day.Charge +  Eve.Calls + Eve.Charge + Night.Calls + Intl.Calls + Intl.Charge  Df Deviance AIC  - Account.Length 1 1009.1 1037.1  - Intl.Mins 1 1009.2 1037.2  - Intl.Charge 1 1009.2 1037.2  - Night.Calls 1 1009.5 1037.5  - Eve.Calls 1 1009.5 1037.5  - Day.Calls 1 1009.6 1037.6  - Eve.Charge 1 1009.6 1037.6  - Eve.Mins 1 1009.6 1037.6  - Night.Mins 1 1010.2 1038.2  <none> 1008.8 1038.8  - Intl.Calls 1 1015.0 1043.0  - VMail.Plan 1 1027.7 1055.7  - Day.Charge 1 1108.0 1136.0  - Int.l.Plan 1 1108.9 1136.9  - CustServ.Calls 1 1151.2 1179.2  Step: AIC=1037.14  Churn ~ Eve.Mins + Night.Mins + Intl.Mins + CustServ.Calls +  Int.l.Plan + VMail.Plan + Day.Calls + Day.Charge + Eve.Calls +  Eve.Charge + Night.Calls + Intl.Calls + Intl.Charge  Df Deviance AIC  - Intl.Mins 1 1009.6 1035.6  - Intl.Charge 1 1009.6 1035.6  - Night.Calls 1 1009.9 1035.9  - Eve.Calls 1 1009.9 1035.9  - Eve.Charge 1 1010.0 1036.0  - Eve.Mins 1 1010.0 1036.0  - Day.Calls 1 1010.0 1036.0  - Night.Mins 1 1010.5 1036.5  <none> 1009.1 1037.1  - Intl.Calls 1 1015.4 1041.4  - VMail.Plan 1 1027.8 1053.8  - Day.Charge 1 1109.0 1135.0  - Int.l.Plan 1 1109.3 1135.3  - CustServ.Calls 1 1151.5 1177.5  Step: AIC=1035.56  Churn ~ Eve.Mins + Night.Mins + CustServ.Calls + Int.l.Plan +  VMail.Plan + Day.Calls + Day.Charge + Eve.Calls + Eve.Charge +  Night.Calls + Intl.Calls + Intl.Charge  Df Deviance AIC  - Night.Calls 1 1010.2 1034.2  - Eve.Calls 1 1010.4 1034.3  - Eve.Charge 1 1010.4 1034.4  - Eve.Mins 1 1010.4 1034.4  - Day.Calls 1 1010.4 1034.4  - Night.Mins 1 1010.9 1034.9  <none> 1009.6 1035.6  - Intl.Calls 1 1015.8 1039.8  - Intl.Charge 1 1017.2 1041.2  - VMail.Plan 1 1028.1 1052.1  - Int.l.Plan 1 1109.3 1133.3  - Day.Charge 1 1112.3 1136.3  - CustServ.Calls 1 1151.7 1175.7  Step: AIC=1034.25  Churn ~ Eve.Mins + Night.Mins + CustServ.Calls + Int.l.Plan +  VMail.Plan + Day.Calls + Day.Charge + Eve.Calls + Eve.Charge +  Intl.Calls + Intl.Charge  Df Deviance AIC  - Eve.Calls 1 1011.0 1033.0  - Eve.Charge 1 1011.1 1033.1  - Eve.Mins 1 1011.1 1033.1  - Day.Calls 1 1011.2 1033.2  - Night.Mins 1 1011.6 1033.7  <none> 1010.2 1034.2  - Intl.Calls 1 1016.6 1038.6  - Intl.Charge 1 1017.6 1039.6  - VMail.Plan 1 1028.5 1050.5  - Int.l.Plan 1 1109.8 1131.8  - Day.Charge 1 1112.9 1134.9  - CustServ.Calls 1 1152.5 1174.5  Step: AIC=1033  Churn ~ Eve.Mins + Night.Mins + CustServ.Calls + Int.l.Plan +  VMail.Plan + Day.Calls + Day.Charge + Eve.Charge + Intl.Calls +  Intl.Charge  Df Deviance AIC  - Eve.Charge 1 1011.7 1031.7  - Eve.Mins 1 1011.7 1031.7  - Day.Calls 1 1011.9 1031.9  - Night.Mins 1 1012.4 1032.4  <none> 1011.0 1033.0  - Intl.Calls 1 1017.5 1037.5  - Intl.Charge 1 1018.3 1038.3  - VMail.Plan 1 1028.8 1048.8  - Int.l.Plan 1 1110.8 1130.8  - Day.Charge 1 1114.7 1134.7  - CustServ.Calls 1 1154.0 1174.0  Step: AIC=1031.73  Churn ~ Eve.Mins + Night.Mins + CustServ.Calls + Int.l.Plan +  VMail.Plan + Day.Calls + Day.Charge + Intl.Calls + Intl.Charge  Df Deviance AIC  - Day.Calls 1 1012.6 1030.6  - Night.Mins 1 1013.1 1031.1  <none> 1011.7 1031.7  - Intl.Calls 1 1018.3 1036.3  - Intl.Charge 1 1018.9 1036.8  - Eve.Mins 1 1029.1 1047.1  - VMail.Plan 1 1029.8 1047.8  - Int.l.Plan 1 1111.1 1129.1  - Day.Charge 1 1115.5 1133.5  - CustServ.Calls 1 1154.3 1172.3  Step: AIC=1030.61  Churn ~ Eve.Mins + Night.Mins + CustServ.Calls + Int.l.Plan +  VMail.Plan + Day.Charge + Intl.Calls + Intl.Charge  Df Deviance AIC  - Night.Mins 1 1014.0 1030.0  <none> 1012.6 1030.6  - Intl.Calls 1 1019.1 1035.1  - Intl.Charge 1 1019.8 1035.8  - Eve.Mins 1 1030.2 1046.2  - VMail.Plan 1 1030.8 1046.8  - Int.l.Plan 1 1111.5 1127.5  - Day.Charge 1 1116.8 1132.8  - CustServ.Calls 1 1154.5 1170.5  Step: AIC=1030.02  Churn ~ Eve.Mins + CustServ.Calls + Int.l.Plan + VMail.Plan +  Day.Charge + Intl.Calls + Intl.Charge  Df Deviance AIC  <none> 1014.0 1030.0  - Intl.Calls 1 1020.7 1034.7  - Intl.Charge 1 1021.0 1035.0  - Eve.Mins 1 1031.8 1045.8  - VMail.Plan 1 1032.7 1046.7  - Int.l.Plan 1 1111.9 1125.9  - Day.Charge 1 1121.3 1135.3  - CustServ.Calls 1 1154.8 1168.8  > formula(x1)  Churn ~ Eve.Mins + CustServ.Calls + Int.l.Plan + VMail.Plan +  Day.Charge + Intl.Calls + Intl.Charge  > glm\_model3<- glm(Churn ~Eve.Mins + CustServ.Calls + Int.l.Plan + VMail.Plan + Day.Charge + Intl.Calls + Intl.Charge,data = train\_data)  > exp(glm\_model3$coefficients)  (Intercept) Eve.Mins CustServ.Calls Int.l.Plan VMail.Plan Day.Charge Intl.Calls Intl.Charge  0.7022931 1.0005957 1.0559019 1.3421714 0.9187070 1.0078609 0.9916826 1.0309707  > pred3 <- predict(glm\_model3,test\_data,type="response")  > plot(data$Churn[-index]~pred3)  > outcome3 <- floor(pred3+0.5)  > ttt3=table(data$Churn[-index],outcome3)  > ttt3  outcome3  0 1  0 563 5  1 89 10   |  | | --- | | > data\_1<-data[data$Churn==1,]  Warning messages:  1: In doTryCatch(return(expr), name, parentenv, handler) :  display list redraw incomplete  2: In doTryCatch(return(expr), name, parentenv, handler) :  invalid graphics state  3: In doTryCatch(return(expr), name, parentenv, handler) :  invalid graphics state  >  > ind\_1<-sample(rownames(data\_1),483)  >  > data\_0<-data[data$Churn==0,]  >  > ind\_0<-sample(rownames(data\_0),483)  > train\_data1<-data[c(ind\_1,ind\_0), -21 ]  >  > table(train\_data1$Churn)  0 1  483 483  > glm\_model4<-glm(Churn ~Eve.Mins + CustServ.Calls + Int.l.Plan + VMail.Plan + Day.Charge + Intl.Calls + Intl.Charge ,family=binomial,data=train\_data1)  > glm\_model4  Call: glm(formula = Churn ~ Eve.Mins + CustServ.Calls + Int.l.Plan +  VMail.Plan + Day.Charge + Intl.Calls + Intl.Charge, family = binomial,  data = train\_data1)  Coefficients:  (Intercept) Eve.Mins CustServ.Calls Int.l.Plan VMail.Plan Day.Charge  -5.441971 0.006684 0.587431 2.094257 -0.783771 0.071874  Intl.Calls Intl.Charge  -0.060476 0.266146  Degrees of Freedom: 965 Total (i.e. Null); 958 Residual  Null Deviance: 1339  Residual Deviance: 1023 AIC: 1039  > summary(glm\_model4)  Call:  glm(formula = Churn ~ Eve.Mins + CustServ.Calls + Int.l.Plan +  VMail.Plan + Day.Charge + Intl.Calls + Intl.Charge, family = binomial,  data = train\_data1)  Deviance Residuals:  Min 1Q Median 3Q Max  -2.76900 -0.83469 -0.05176 0.87634 2.41004  Coefficients:  Estimate Std. Error z value Pr(>|z|)  (Intercept) -5.441971 0.573201 -9.494 < 2e-16 \*\*\*  Eve.Mins 0.006684 0.001543 4.332 1.48e-05 \*\*\*  CustServ.Calls 0.587431 0.055799 10.528 < 2e-16 \*\*\*  Int.l.Plan 2.094257 0.230233 9.096 < 2e-16 \*\*\*  VMail.Plan -0.783771 0.188956 -4.148 3.36e-05 \*\*\*  Day.Charge 0.071874 0.007916 9.080 < 2e-16 \*\*\*  Intl.Calls -0.060476 0.030771 -1.965 0.04937 \*  Intl.Charge 0.266146 0.100780 2.641 0.00827 \*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for binomial family taken to be 1)  Null deviance: 1339.2 on 965 degrees of freedom  Residual deviance: 1023.5 on 958 degrees of freedom  AIC: 1039.5  Number of Fisher Scoring iterations: 4  > pred4 <- predict(glm\_model4,test\_data,type="response")  > outcome4 <- floor(pred4+0.5)  > table(outcome4)  outcome4  0 1  478 189  > ttt4<- table(data$Churn[-index],outcome4)  > ttt4  outcome4  0 1  0 451 117   1. 27 72   > exp(cbind(Odds\_and\_OR=coef(glm\_model4),confint(glm\_model4)))  Waiting for profiling to be done...  Odds\_and\_OR 2.5 % 97.5 %  (Intercept) 0.004330938 0.001374368 0.01302715  Eve.Mins 1.006705971 1.003691442 1.00978589  CustServ.Calls 1.799360044 1.616709108 2.01238490  Int.l.Plan 8.119403312 5.230140342 12.91663356  VMail.Plan 0.456680467 0.313961751 0.65911549  Day.Charge 1.074520278 1.058271928 1.09165809  Intl.Calls 0.941316223 0.885754238 0.99961207  Intl.Charge 1.304925154 1.072221833 1.59240956 | |  | | |  | | --- | |  | |   ##accuracy##  > acc<-(ttt4[1]+ttt4[4])/(ttt4[1]+ttt4[3]+ttt4[2]+ttt4[4])  > acc  [1] 0.7841079  >  > sens=ttt4[1]/(ttt4[1]+ttt4[3])  > sens  [1] 0.7940141  >  > spec=ttt4[4]/(ttt4[4]+ttt4[2])  > spec  [1] 0.7272727  install.packages("caret")  library(caret)  > confusionMatrix(ttt4)  Confusion Matrix and Statistics  outcome4  0 1  0 447 121  1 26 73    Accuracy : 0.7796  95% CI : (0.7462, 0.8105)  No Information Rate : 0.7091  P-Value [Acc > NIR] : 2.387e-05    Kappa : 0.3756  Mcnemar's Test P-Value : 8.975e-15    Sensitivity : 0.9450  Specificity : 0.3763  Pos Pred Value : 0.7870  Neg Pred Value : 0.7374  Prevalence : 0.7091  Detection Rate : 0.6702  Detection Prevalence : 0.8516  Balanced Accuracy : 0.6607    'Positive' Class : 0   |  | | --- | | > pred5<-predict(glm\_model4,train\_data1,type = "response")  > pred6<-prediction(predictions = pred5,labels = train\_data1$Churn )  > eval<- performance(pred6,measure = "tpr", x.measure = "fpr")  > plot(eval,colorize=T, main="ROC curve",col="blue",lwd=5)  > auc<-performance(pred6,measure = "auc")  > str(auc)  Formal class 'performance' [package "ROCR"] with 6 slots  ..@ x.name : chr "None"  ..@ y.name : chr "Area under the ROC curve"  ..@ alpha.name : chr "none"  ..@ x.values : list()  ..@ y.values :List of 1  .. ..$ : num 0.821  ..@ alpha.values: list()  > as.numeric(auc@y.values) #Printing auc value  [1] 0.8211703 | |  | | |  | | --- | |  | |   ################################################################### |