

```
$dataFile
```

```
[1] "payer_model_DA_GP&iOS_mkt_2019-01-01_2019-03-31.rds"
```

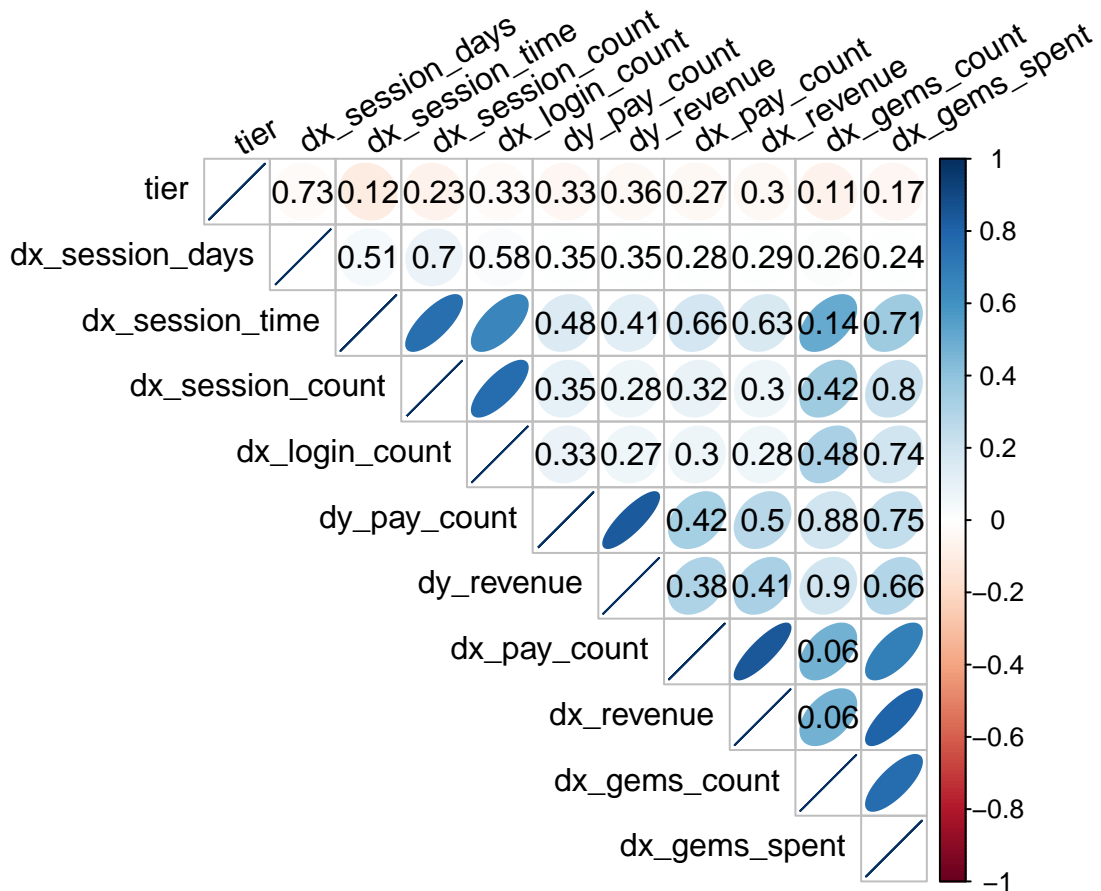
```
$strainRegDate
```

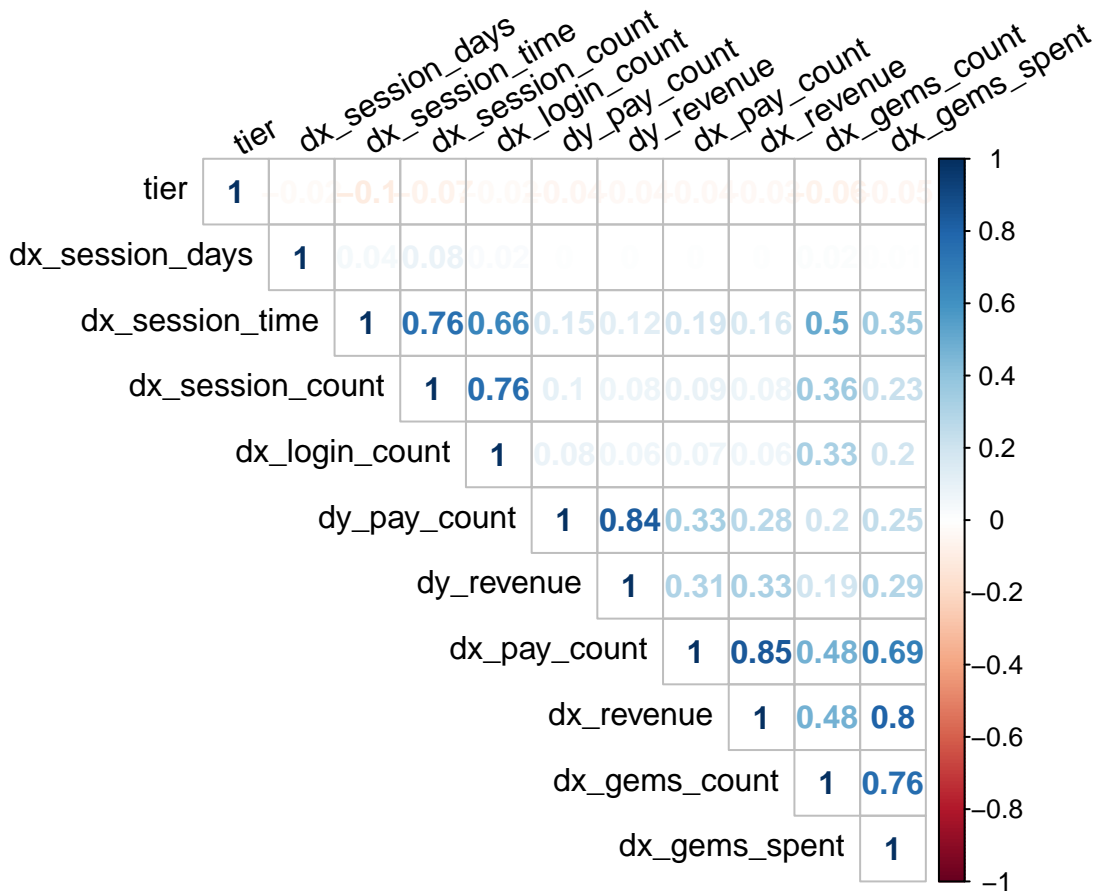
```
[1] "2019-01-01" "2019-01-31"
```

```
$testRegDate
```

```
[1] "2019-03-01" "2019-03-31"
```

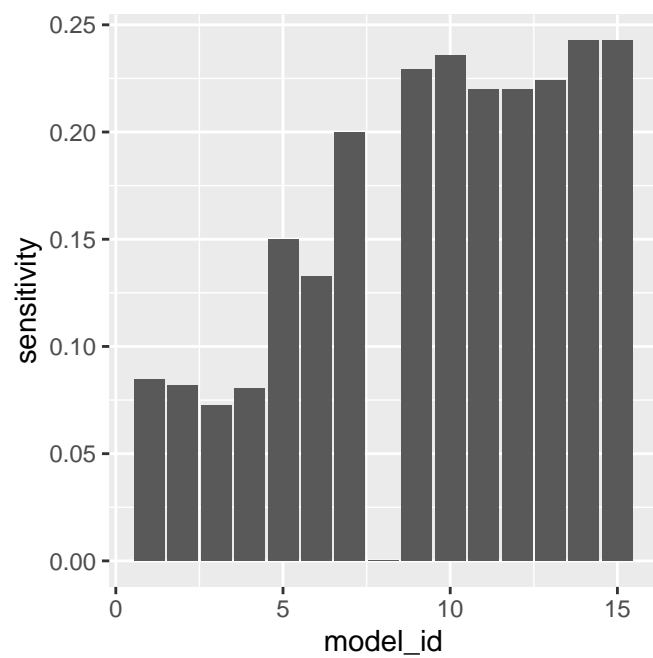
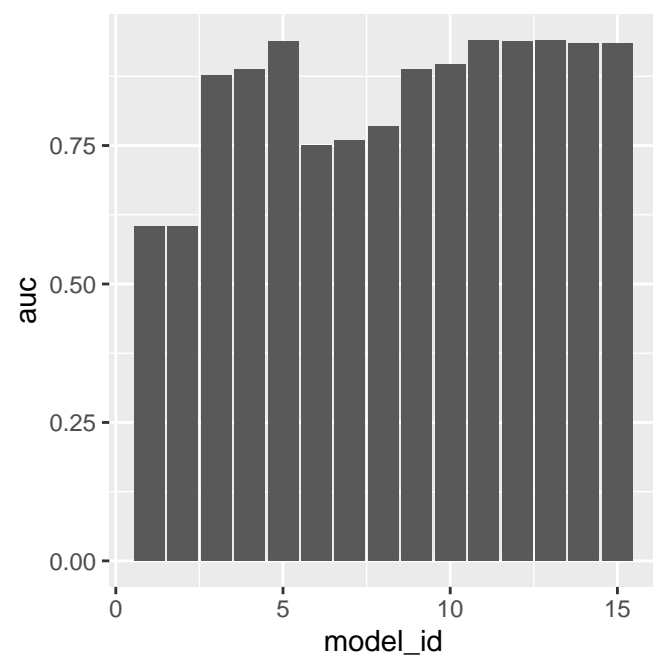
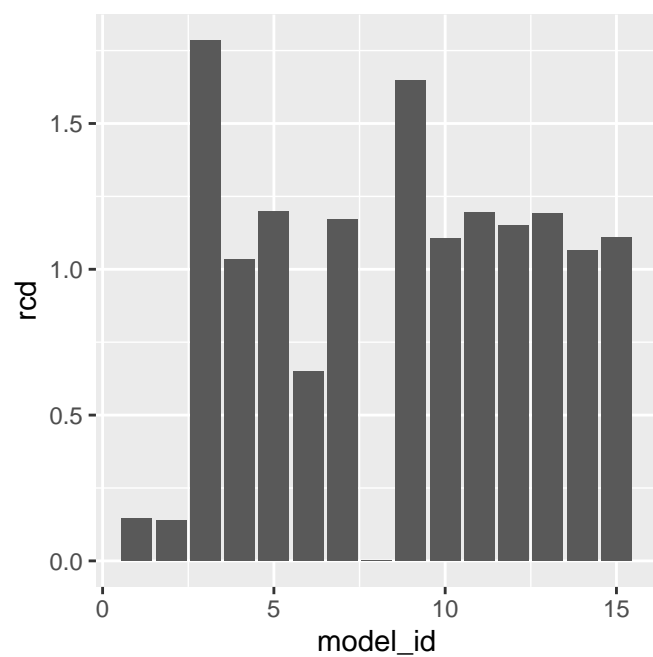
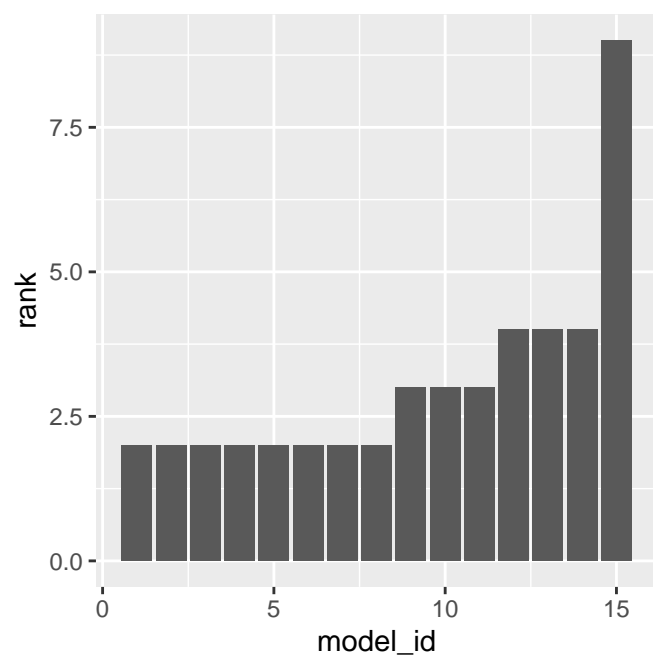
```
[1] "0_data/payer_model_DA_GP&iOS_mkt_2019-01-01_2019-03-31.rds"
```





	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m15
dx_revenue	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1
dx_pay_count	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
dx_login_count	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1
dx_session_count	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1
dx_session_time	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1
dx_gems_count	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1
dx_gems_spent	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1
tier	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1

	model_id	rank	auc	cut_off	rcd	sensitivity
1	1	2	0.6036306	0.23614148	0.1474531	0.08445040
2	2	2	0.6036291	0.85416807	0.1394102	0.08176944
3	3	2	0.8757994	0.03451026	1.7841823	0.07238606
4	4	2	0.8874687	0.05006490	1.0335121	0.08042895
5	5	2	0.9381342	0.07682594	1.1997319	0.15013405
6	6	2	0.7507329	0.02133208	0.6487936	0.13270777
7	7	2	0.7595529	0.00461774	1.1715818	0.19973190
8	8	2	0.7849696	0.99993389	0.0000000	0.00000000
9	9	3	0.8868107	0.03451053	1.6461126	0.22922252
10	10	3	0.8957020	0.04449531	1.1045576	0.23592493
11	11	3	0.9393377	0.06617483	1.1943700	0.21983914
12	12	4	0.9381600	0.06905214	1.1514745	0.21983914
13	13	4	0.9394445	0.06633997	1.1903485	0.22386059
14	14	4	0.9336670	0.08370598	1.0656836	0.24262735
15	15	9	0.9343646	0.08830313	1.1085791	0.24262735



THIS SUMMARY IS OUTDATED

WE FOUND OUT THAT WE NEED CROSS-VALIDATION TO GET ROBUST RESULTS

Summary for DA:

Based on correlation matrix, these variables should not be together

dx_session_time, dx_session_count, dx_login_count

dx_pay_count, dx_revenue

dx_gems_spent, dx_gems_count

dx_gems_spent, dx_revenue

dx_gems_spent, dx_pay_count

We will test 1-variable models

dx_revenue ~ dx_pay_count (we choose revenue)

dx_session_time > dx_session_count > dx_login_count (as expected)

dx_gems_count > dx_gems_spent

Check session features with dx_revenue

dx_session_count ~> dx_session_time > dx_login_count (time is cont.)

Check gem features with previous

dx_gems_spent = dx_gems_count (spent is cont.)

gems do not add anything and are correlated with revenue

Check tier

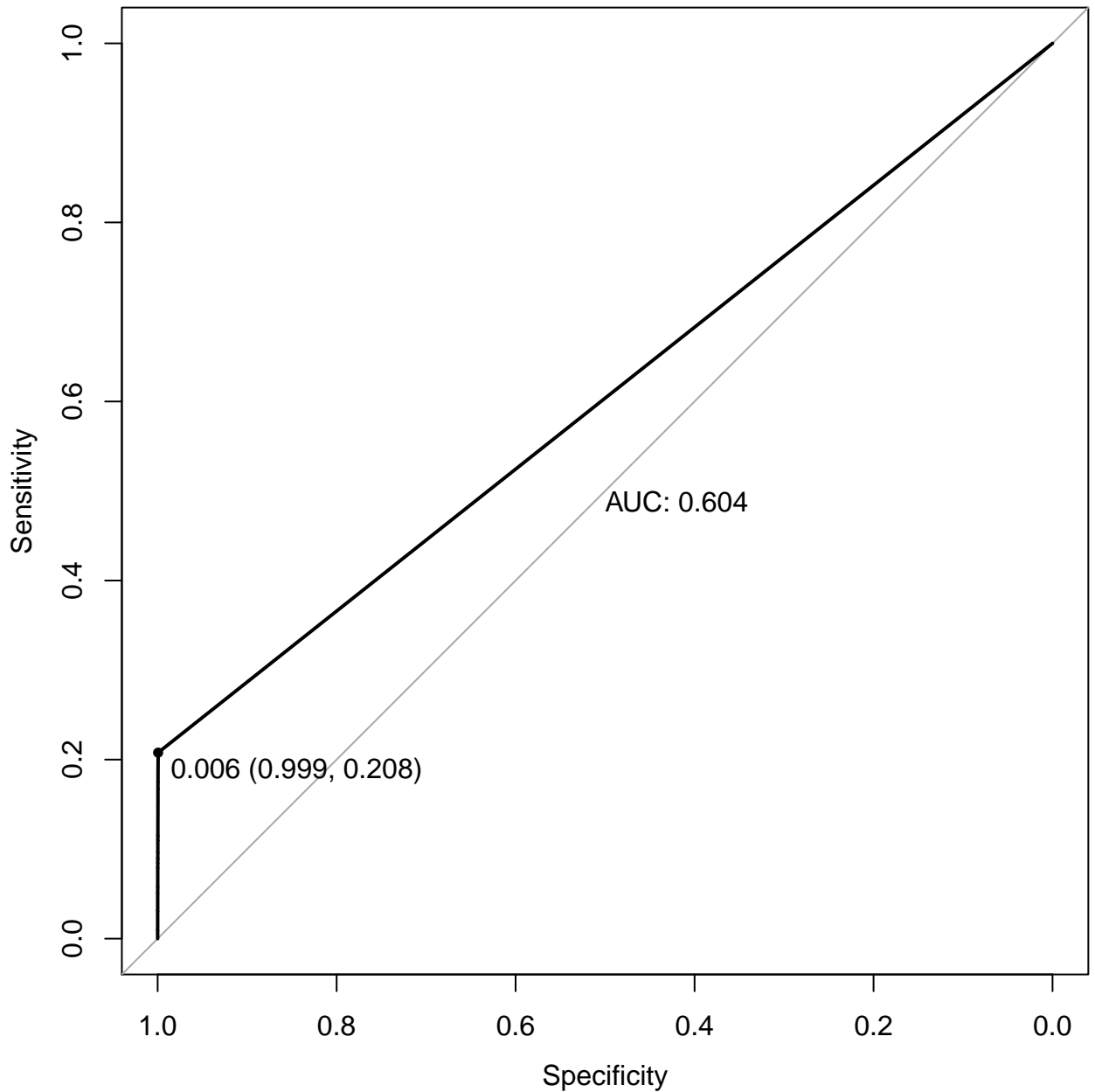
it bends the ROC curve in a weird way but improves the model

Comparison with the full model shows that it has slightly better performance.

It should be compared with cross-validation whether this difference is robust.

Best model so far has: dx_revenue, dx_session_time, tier

ROC curve for model 1



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.0593  -0.0593  -0.0593   3.5623

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.34325    0.03369 -188.29  <2e-16 ***
dx_revenue   0.71466    0.02207   32.38  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 13144  on 501486  degrees of freedom
AIC: 13148

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.6036
```

```
$relativeCountDifference
```

```
[1] 0.1474531
```

```
$optimal_cut_off
```

```
[1] 0.2361415
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338771	683
TRUE	47	63

Accuracy : 0.9979

95% CI : (0.9977, 0.998)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 0.2864

Kappa : 0.1467

McNemar's Test P-Value : <2e-16

Sensitivity : 0.0844504

Specificity : 0.9998613

Pos Pred Value : 0.5727273

Neg Pred Value : 0.9979879

Prevalence : 0.0021969

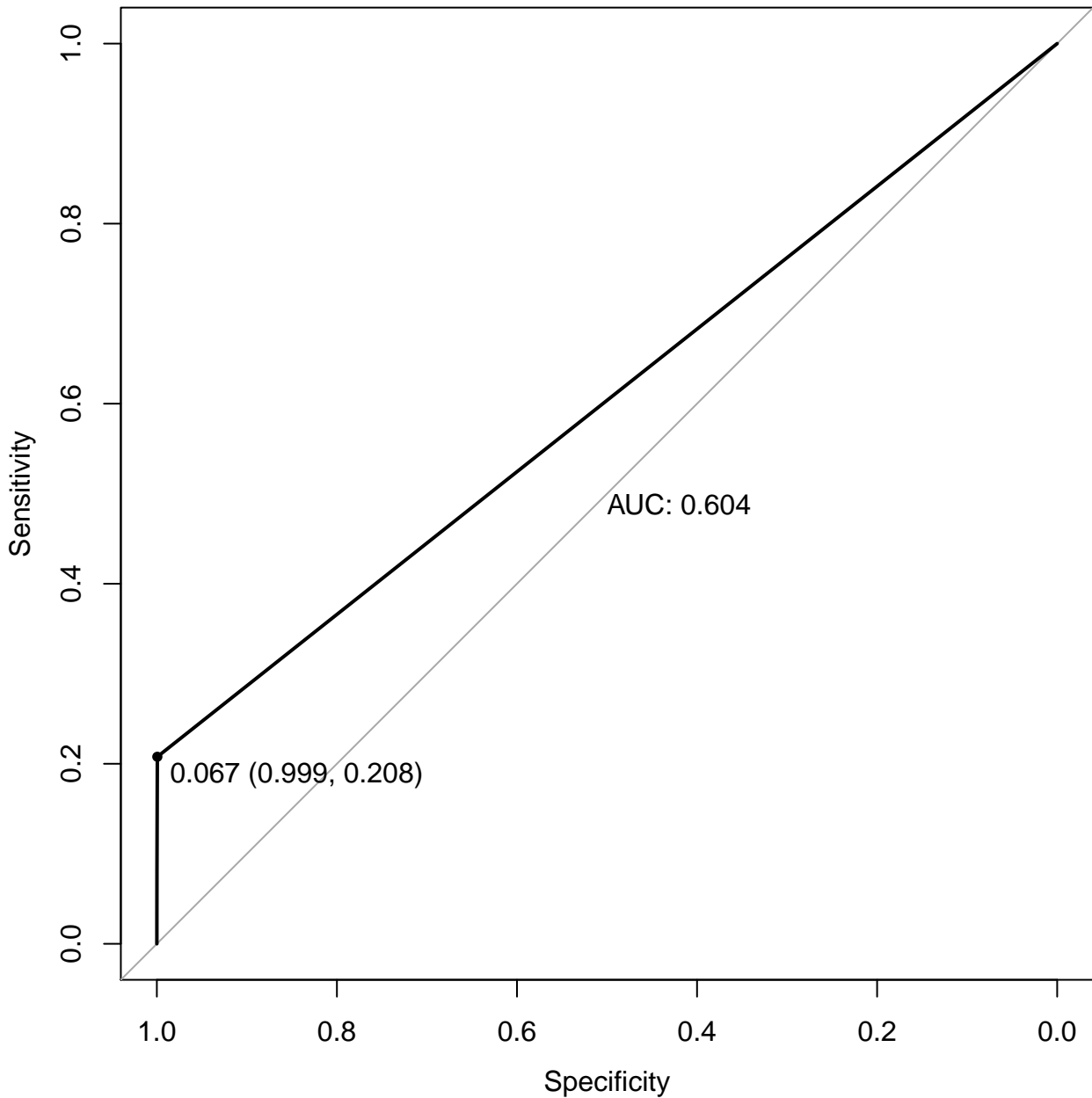
Detection Rate : 0.0001855

Detection Prevalence : 0.0003239

Balanced Accuracy : 0.5421558

'Positive' Class : TRUE

ROC curve for model 2



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.6698  -0.0583  -0.0583  -0.0583   3.5718

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.37725     0.03426 -186.16  <2e-16 ***
dx_pay_count   4.49014     0.13562   33.11  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 12828  on 501486  degrees of freedom
AIC: 12832

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.6036
```

```
$relativeCountDifference
```

```
[1] 0.1394102
```



```
$optimal_cut_off
```

```
[1] 0.8541681
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338775	685
TRUE	43	61

Accuracy : 0.9979

95% CI : (0.9977, 0.998)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 0.2618

Kappa : 0.1431

Mcnemar's Test P-Value : <2e-16

Sensitivity : 0.0817694

Specificity : 0.9998731

Pos Pred Value : 0.5865385

Neg Pred Value : 0.9979821

Prevalence : 0.0021969

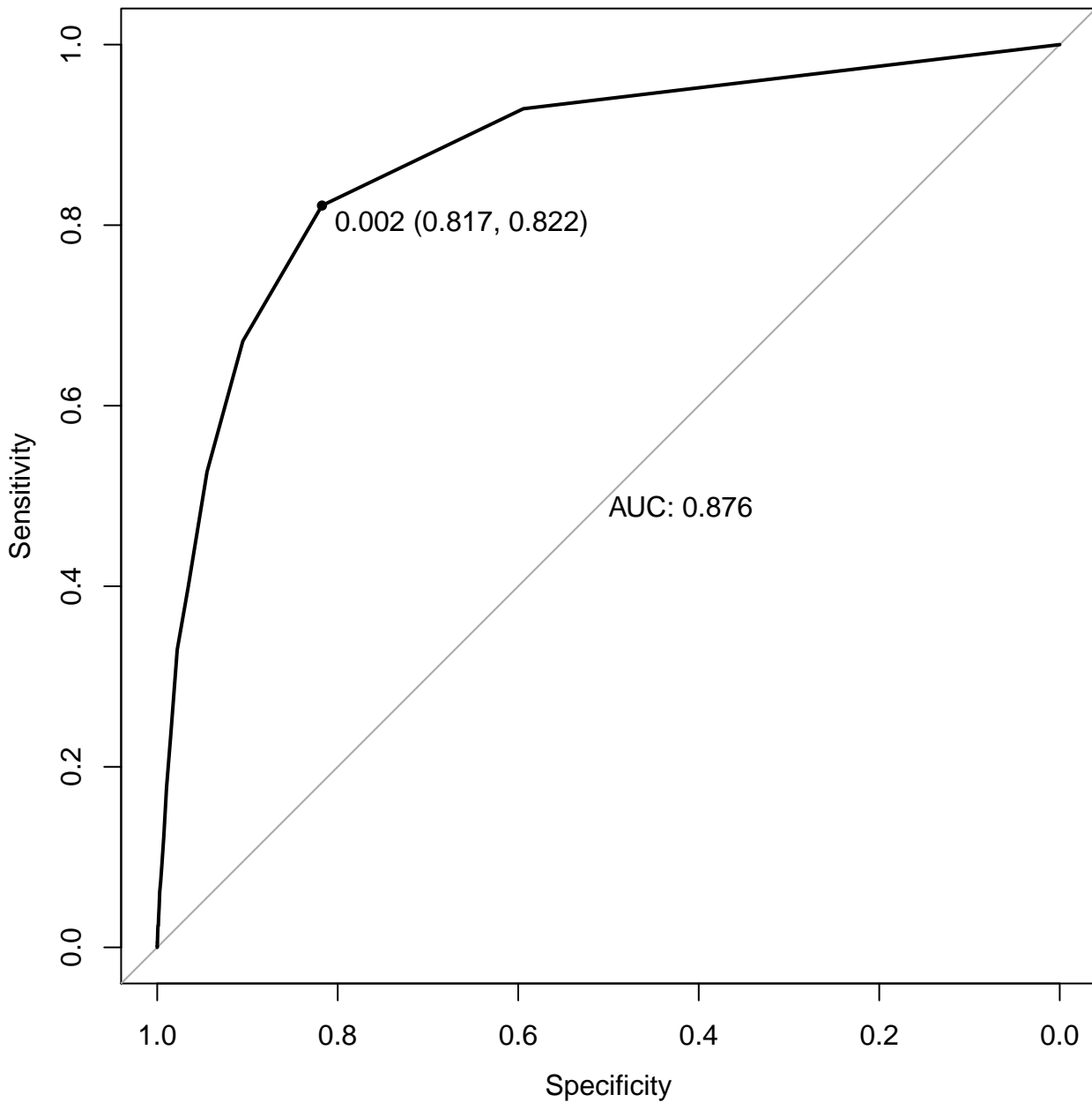
Detection Rate : 0.0001796

Detection Prevalence : 0.0003063

Balanced Accuracy : 0.5408213

'Positive' Class : TRUE

ROC curve for model 3



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.4929  -0.0606  -0.0520  -0.0520   3.6349

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.90911    0.04052  -170.49  <2e-16 ***
dx_login_count  0.30422    0.00682   44.61  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 13346  on 501486  degrees of freedom
AIC: 13350

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.8758
```

```
$relativeCountDifference
```

```
[1] 1.784182
```

```
$optimal_cut_off
```

```
[1] 0.03451026
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	337541	692
TRUE	1277	54

Accuracy : 0.9942
95% CI : (0.9939, 0.9945)
No Information Rate : 0.9978
P-Value [Acc > NIR] : 1

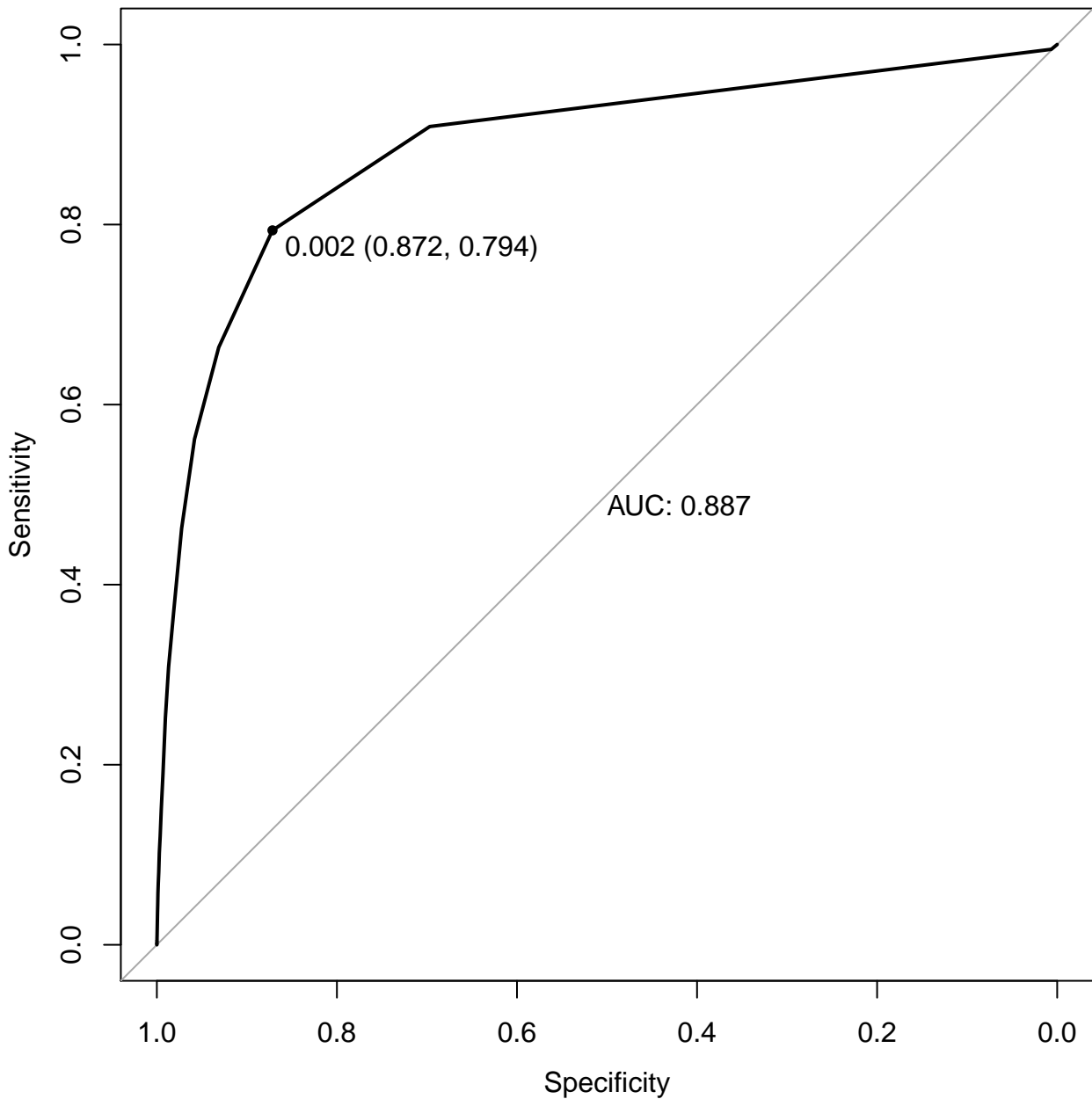
Kappa : 0.0493

McNemar's Test P-Value : <2e-16

Sensitivity : 0.072386
Specificity : 0.996231
Pos Pred Value : 0.040571
Neg Pred Value : 0.997954
Prevalence : 0.002197
Detection Rate : 0.000159
Detection Prevalence : 0.003920
Balanced Accuracy : 0.534309

'Positive' Class : TRUE

ROC curve for model 4




```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2550  -0.0588  -0.0506  -0.0506   3.7307

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.958251   0.040687 -171.02  <2e-16 ***
dx_session_count  0.298802   0.005471   54.62  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 12709  on 501486  degrees of freedom
AIC: 12713

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.8875
```

```
$relativeCountDifference
```

```
[1] 1.033512
```

```
$optimal_cut_off
```

```
[1] 0.0500649
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338107	686
TRUE	711	60

Accuracy : 0.9959

95% CI : (0.9957, 0.9961)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 1.0000

Kappa : 0.077

McNemar's Test P-Value : 0.5208

Sensitivity : 0.0804290

Specificity : 0.9979015

Pos Pred Value : 0.0778210

Neg Pred Value : 0.9979752

Prevalence : 0.0021969

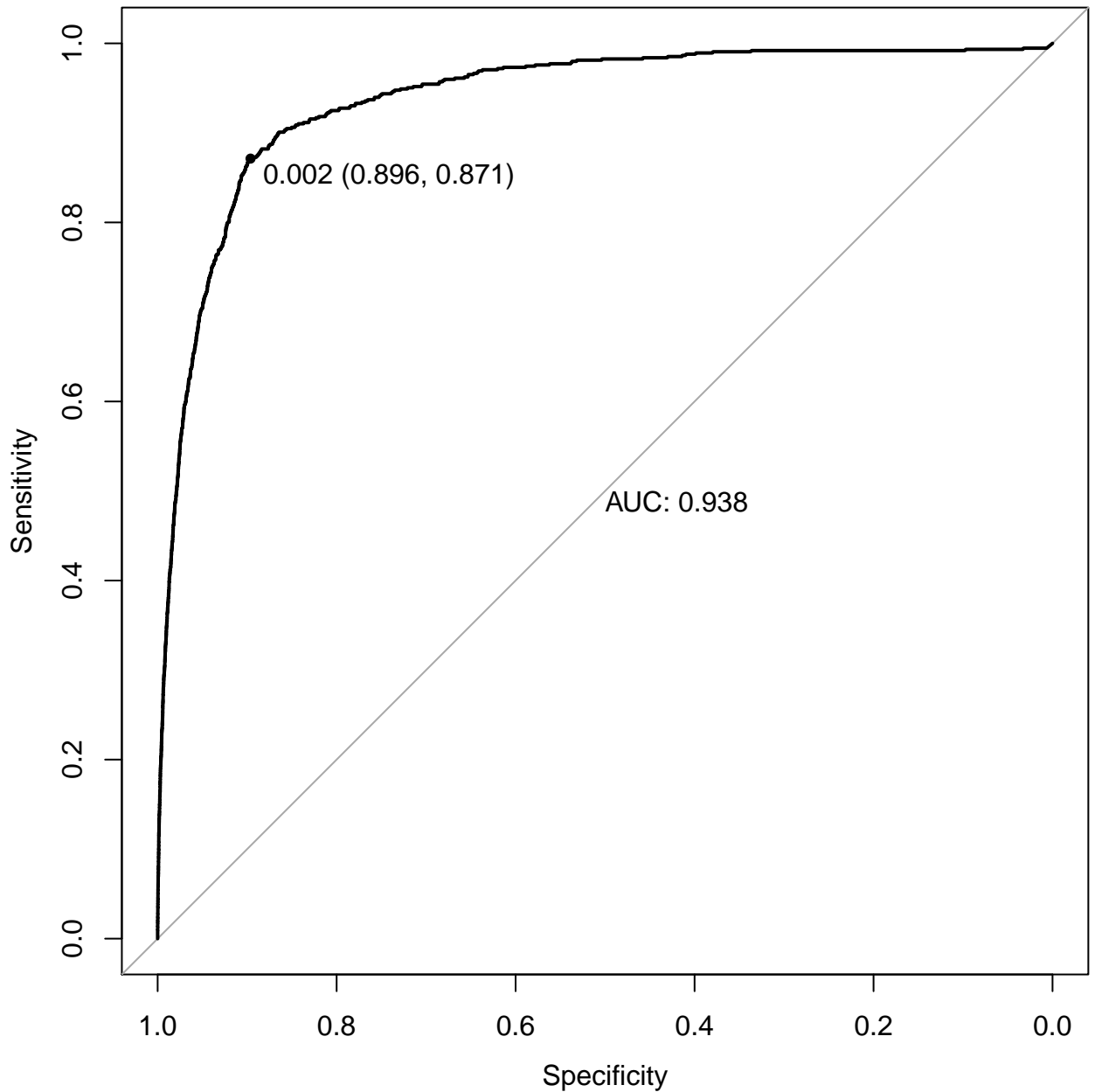
Detection Rate : 0.0001767

Detection Prevalence : 0.0022706

Balanced Accuracy : 0.5391652

'Positive' Class : TRUE

ROC curve for model 5



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.2592  -0.0487  -0.0448  -0.0433   3.7463

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -7.016658    0.043097  -162.81  <2e-16 ***
dx_session_time  0.015916    0.000239   66.58  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 11360  on 501486  degrees of freedom
AIC: 11364

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.9381
```

```
$relativeCountDifference
```

```
[1] 1.199732
```

```
$optimal_cut_off
```

```
[1] 0.07682594
```


\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338035	634
TRUE	783	112

Accuracy : 0.9958

95% CI : (0.9956, 0.996)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 1

Kappa : 0.1344

McNemar's Test P-Value : 8.436e-05

Sensitivity : 0.1501340

Specificity : 0.9976890

Pos Pred Value : 0.1251397

Neg Pred Value : 0.9981280

Prevalence : 0.0021969

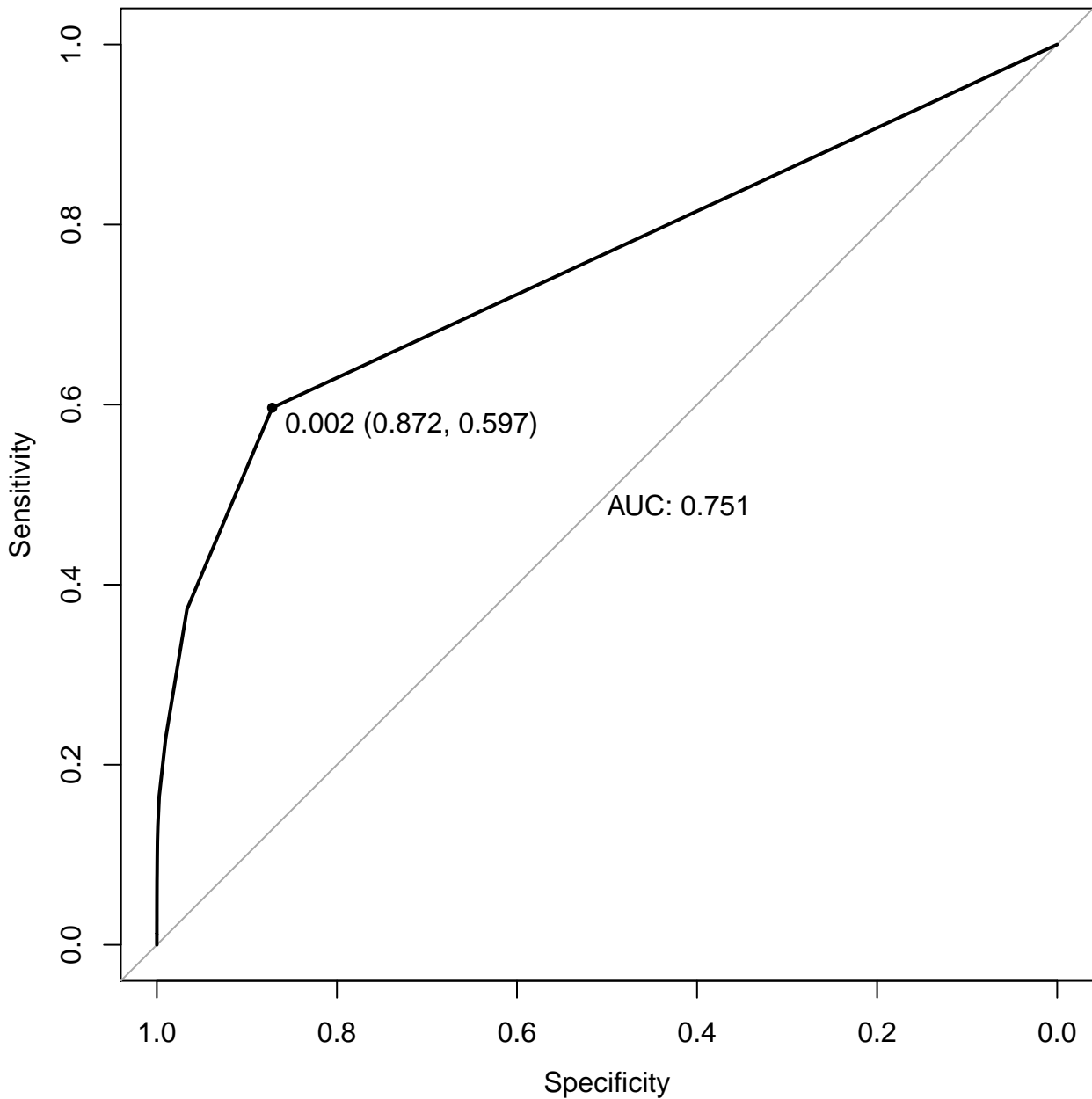
Detection Rate : 0.0003298

Detection Prevalence : 0.0026357

Balanced Accuracy : 0.5739115

'Positive' Class : TRUE

ROC curve for model 6



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.9156  -0.0544  -0.0544  -0.0544   3.6101

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.51510    0.03585  -181.7  <2e-16 ***
dx_gems_count  0.66147    0.01602   41.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 12779  on 501486  degrees of freedom
AIC: 12783

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.7507
```

```
$relativeCountDifference
```

```
[1] 0.6487936
```

```
$optimal_cut_off
```

```
[1] 0.02133208
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338433	647
TRUE	385	99

Accuracy : 0.997

95% CI : (0.9968, 0.9971)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 1

Kappa : 0.1595

McNemar's Test P-Value : 4.489e-16

Sensitivity : 0.1327078

Specificity : 0.9988637

Pos Pred Value : 0.2045455

Neg Pred Value : 0.9980919

Prevalence : 0.0021969

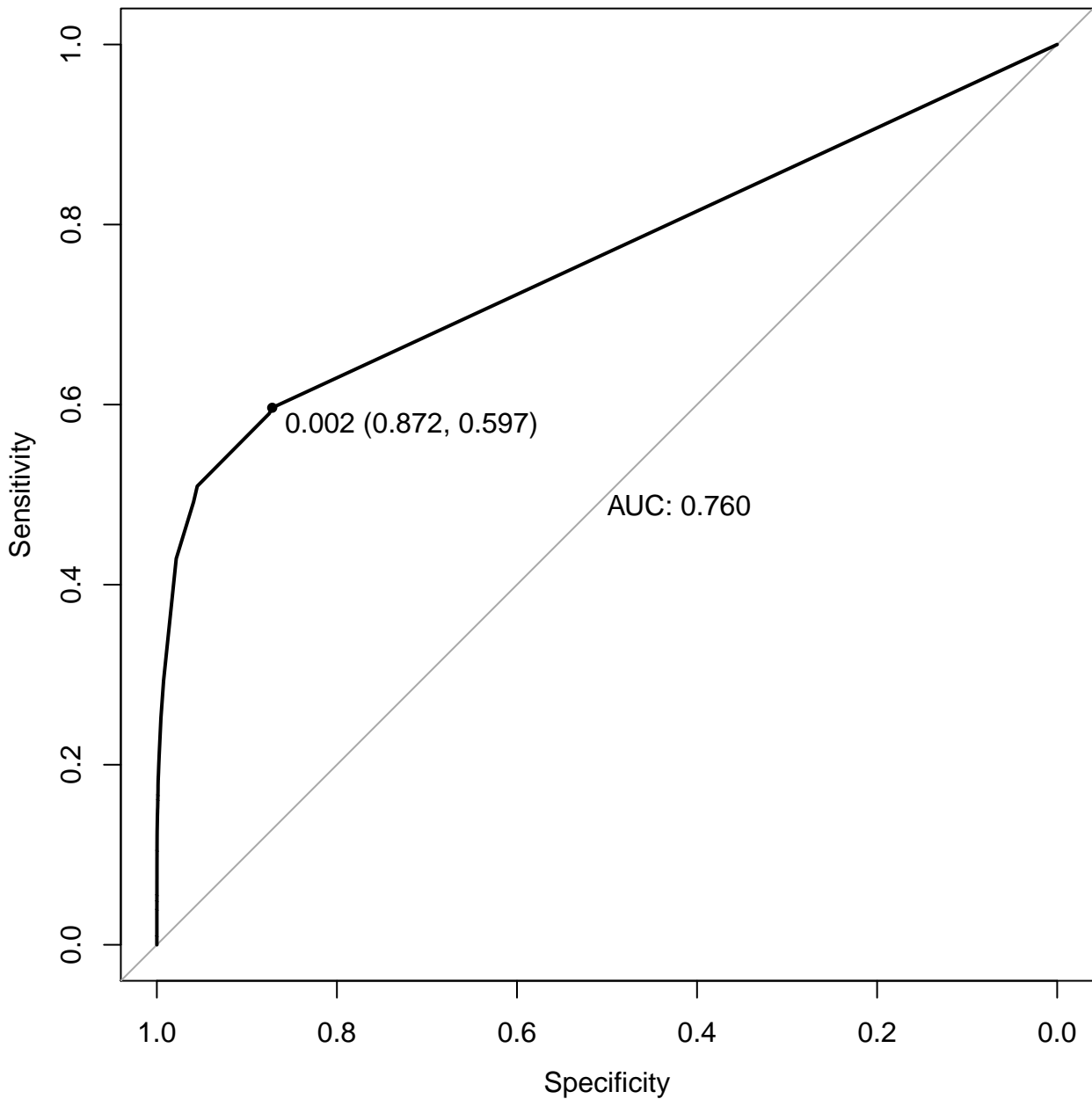
Detection Rate : 0.0002916

Detection Prevalence : 0.0014254

Balanced Accuracy : 0.5657857

'Positive' Class : TRUE

ROC curve for model 7



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.0578  -0.0578  -0.0578   3.5763

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.3933974  0.0341619 -187.15  <2e-16 ***
dx_gems_spent  0.0252408  0.0007699   32.78  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 13028  on 501486  degrees of freedom
AIC: 13032

Number of Fisher Scoring iterations: 9

```



```
$auc
```

```
Area under the curve: 0.7596
```

```
$relativeCountDifference
```

```
[1] 1.171582
```

```
$optimal_cut_off
```

```
[1] 0.00461774
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338093	597
TRUE	725	149

Accuracy : 0.9961

95% CI : (0.9959, 0.9963)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 1.0000000

Kappa : 0.182

McNemar's Test P-Value : 0.0004778

Sensitivity : 0.1997319

Specificity : 0.9978602

Pos Pred Value : 0.1704805

Neg Pred Value : 0.9982373

Prevalence : 0.0021969

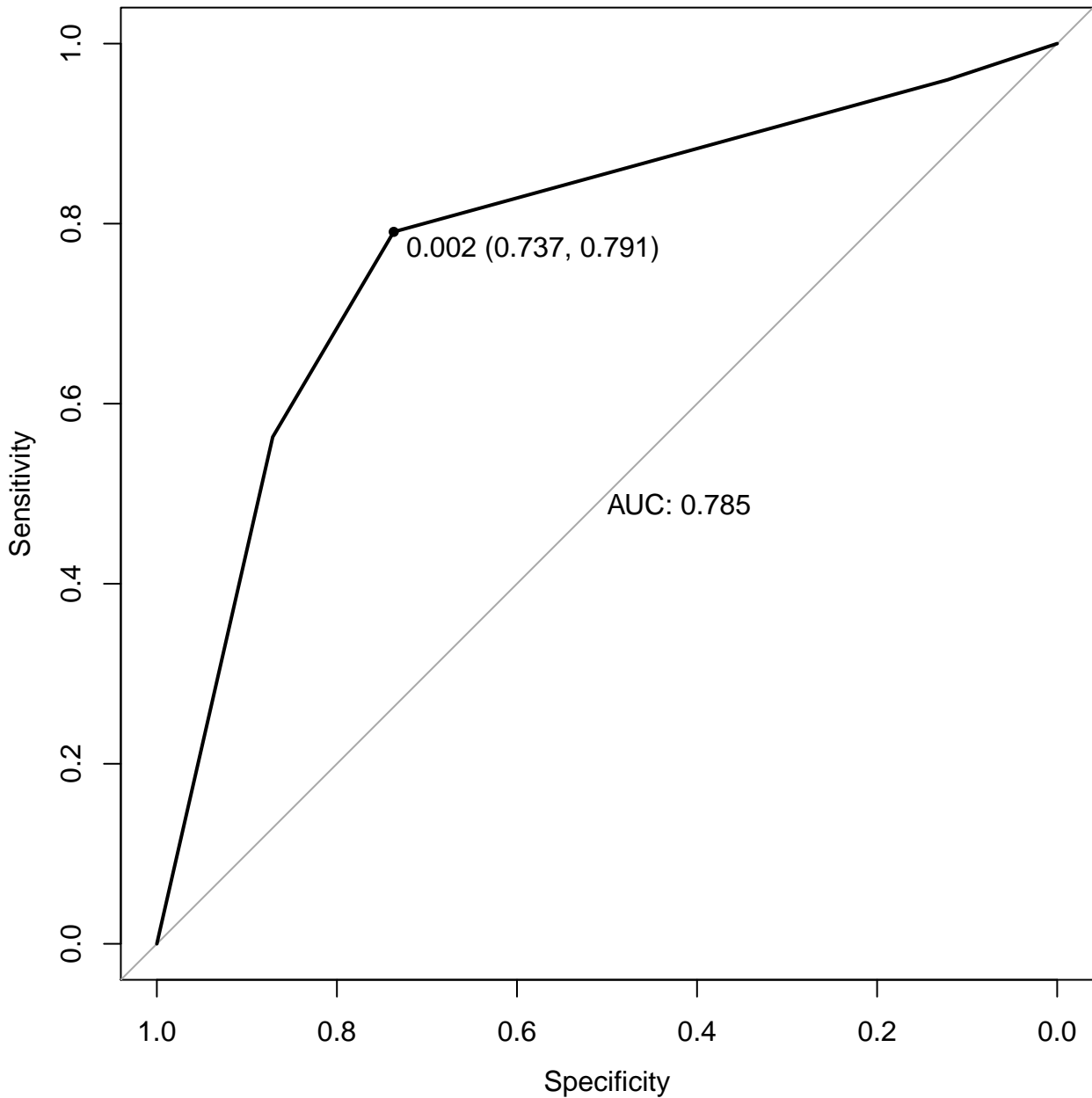
Detection Rate : 0.0004388

Detection Prevalence : 0.0025739

Balanced Accuracy : 0.5987961

'Positive' Class : TRUE

ROC curve for model 8



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.1422  -0.0385  -0.0385  -0.0385   4.1264

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.28085    0.06540  -50.16  <2e-16 ***
tier         -1.30813    0.03538  -36.97  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 13043  on 501486  degrees of freedom
AIC: 13047

Number of Fisher Scoring iterations: 10

```

```
$auc
```

```
Area under the curve: 0.785
```

```
$relativeCountDifference
```

```
[1] 0
```

```
$optimal_cut_off
```

```
[1] 0.9999339
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338818	746
TRUE	0	0

Accuracy : 0.9978

95% CI : (0.9976, 0.998)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 0.5097

Kappa : 0

Mcnemar's Test P-Value : <2e-16

Sensitivity : 0.000000

Specificity : 1.000000

Pos Pred Value : NaN

Neg Pred Value : 0.997803

Prevalence : 0.002197

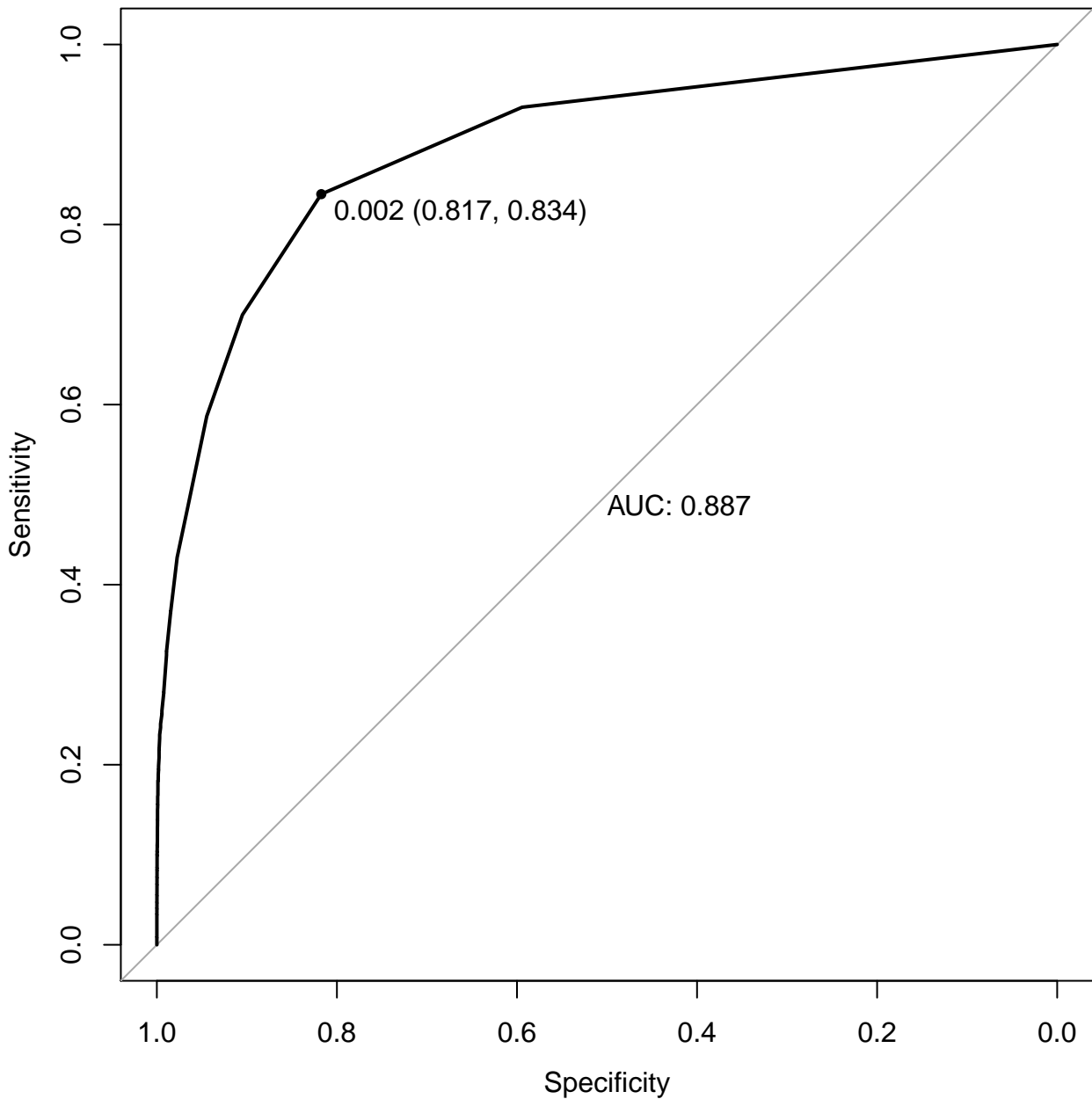
Detection Rate : 0.000000

Detection Prevalence : 0.000000

Balanced Accuracy : 0.500000

'Positive' Class : TRUE

ROC curve for model 9



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.0569  -0.0493  -0.0493   3.6639

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.995263   0.042695 -163.84  <2e-16 ***
dx_revenue     0.592931   0.019799   29.95  <2e-16 ***
dx_login_count 0.284265   0.007255   39.18  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 12160  on 501485  degrees of freedom
AIC: 12166

Number of Fisher Scoring iterations: 10

```

```
$auc
```

```
Area under the curve: 0.8868
```

```
$relativeCountDifference
```

```
[1] 1.646113
```

```
$optimal_cut_off
```

```
[1] 0.03451053
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	337761	575
TRUE	1057	171

Accuracy : 0.9952
95% CI : (0.995, 0.9954)
No Information Rate : 0.9978
P-Value [Acc > NIR] : 1

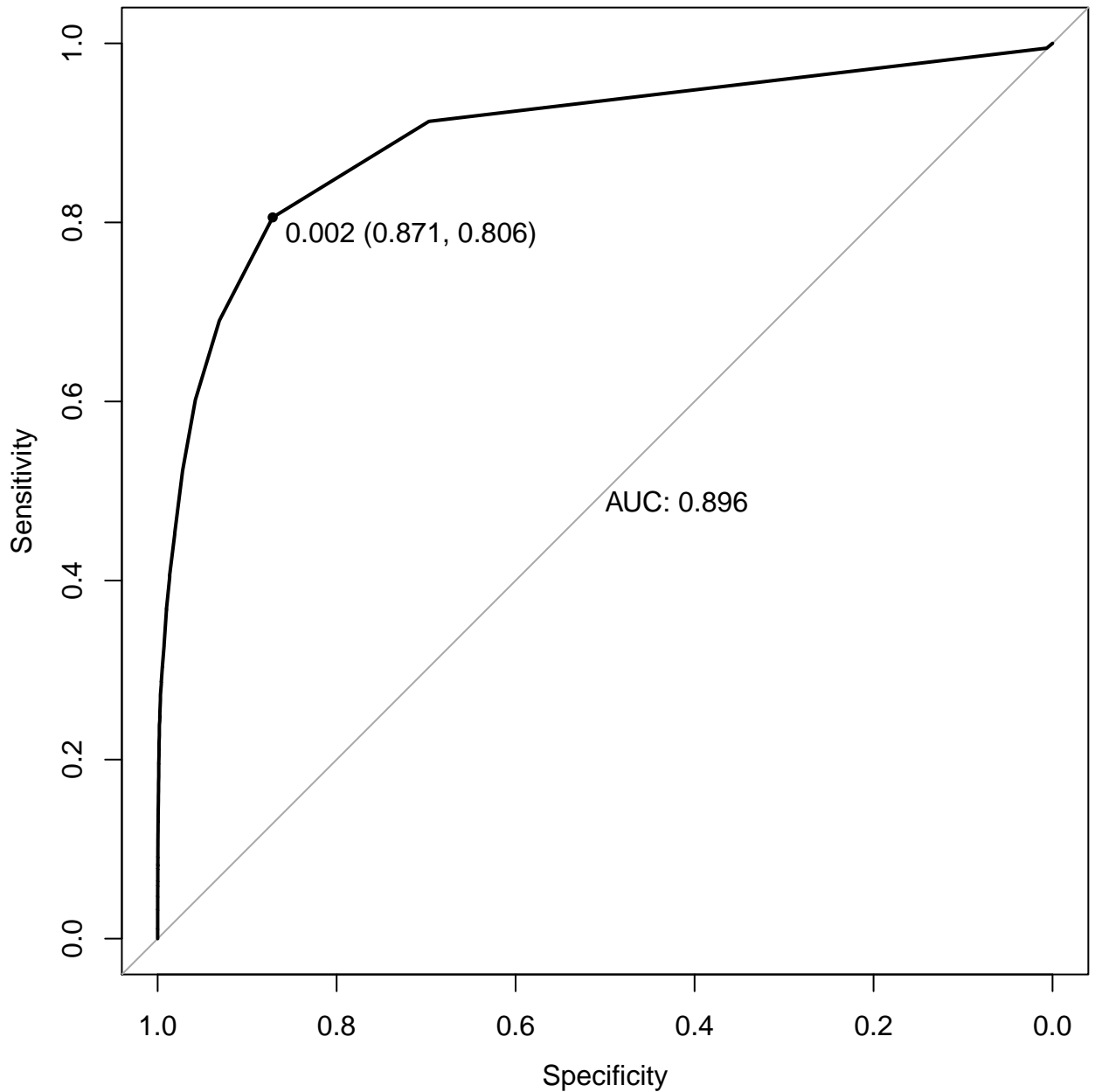
Kappa : 0.171

McNemar's Test P-Value : <2e-16

Sensitivity : 0.2292225
Specificity : 0.9968803
Pos Pred Value : 0.1392508
Neg Pred Value : 0.9983005
Prevalence : 0.0021969
Detection Rate : 0.0005036
Detection Prevalence : 0.0036164
Balanced Accuracy : 0.6130514

'Positive' Class : TRUE

ROC curve for model 10



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8.4904  -0.0555  -0.0483  -0.0483   3.7503

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -7.031327   0.042615  -165.00  <2e-16 ***
dx_revenue      0.536702   0.019040   28.19  <2e-16 ***
dx_session_count 0.278517   0.005794   48.07  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 11681  on 501485  degrees of freedom
AIC: 11687

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.8957
```

```
$relativeCountDifference
```

```
[1] 1.104558
```



```
$optimal_cut_off
```

```
[1] 0.04449531
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338170	570
TRUE	648	176

Accuracy : 0.9964

95% CI : (0.9962, 0.9966)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 1.00000

Kappa : 0.2224

McNemar's Test P-Value : 0.02736

Sensitivity : 0.2359249

Specificity : 0.9980875

Pos Pred Value : 0.2135922

Neg Pred Value : 0.9983173

Prevalence : 0.0021969

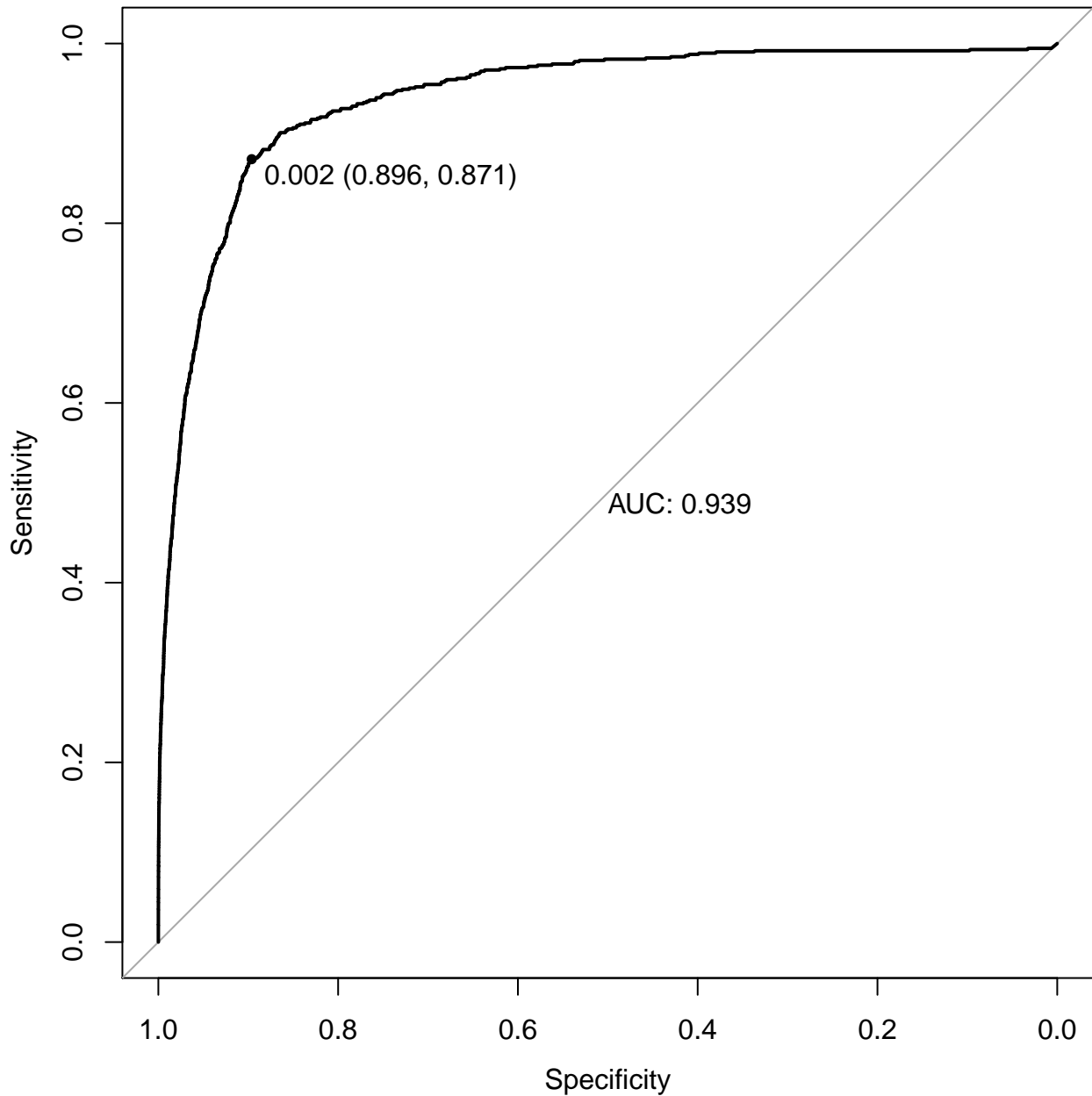
Detection Rate : 0.0005183

Detection Prevalence : 0.0024266

Balanced Accuracy : 0.6170062

'Positive' Class : TRUE

ROC curve for model 11



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.4636  -0.0483  -0.0448  -0.0434   3.7440

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -7.0078693   0.0434432  -161.31  <2e-16 ***
dx_revenue     0.3409664   0.0189257   18.02  <2e-16 ***
dx_session_time 0.0146383   0.0002586   56.60  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 10969  on 501485  degrees of freedom
AIC: 10975

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.9393
```

```
$relativeCountDifference
```

```
[1] 1.19437
```

```
$optimal_cut_off
```

```
[1] 0.06617483
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338091	582
TRUE	727	164

Accuracy : 0.9961

95% CI : (0.9959, 0.9964)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 1

Kappa : 0.1984

McNemar's Test P-Value : 6.889e-05

Sensitivity : 0.219839

Specificity : 0.997854

Pos Pred Value : 0.184063

Neg Pred Value : 0.998282

Prevalence : 0.002197

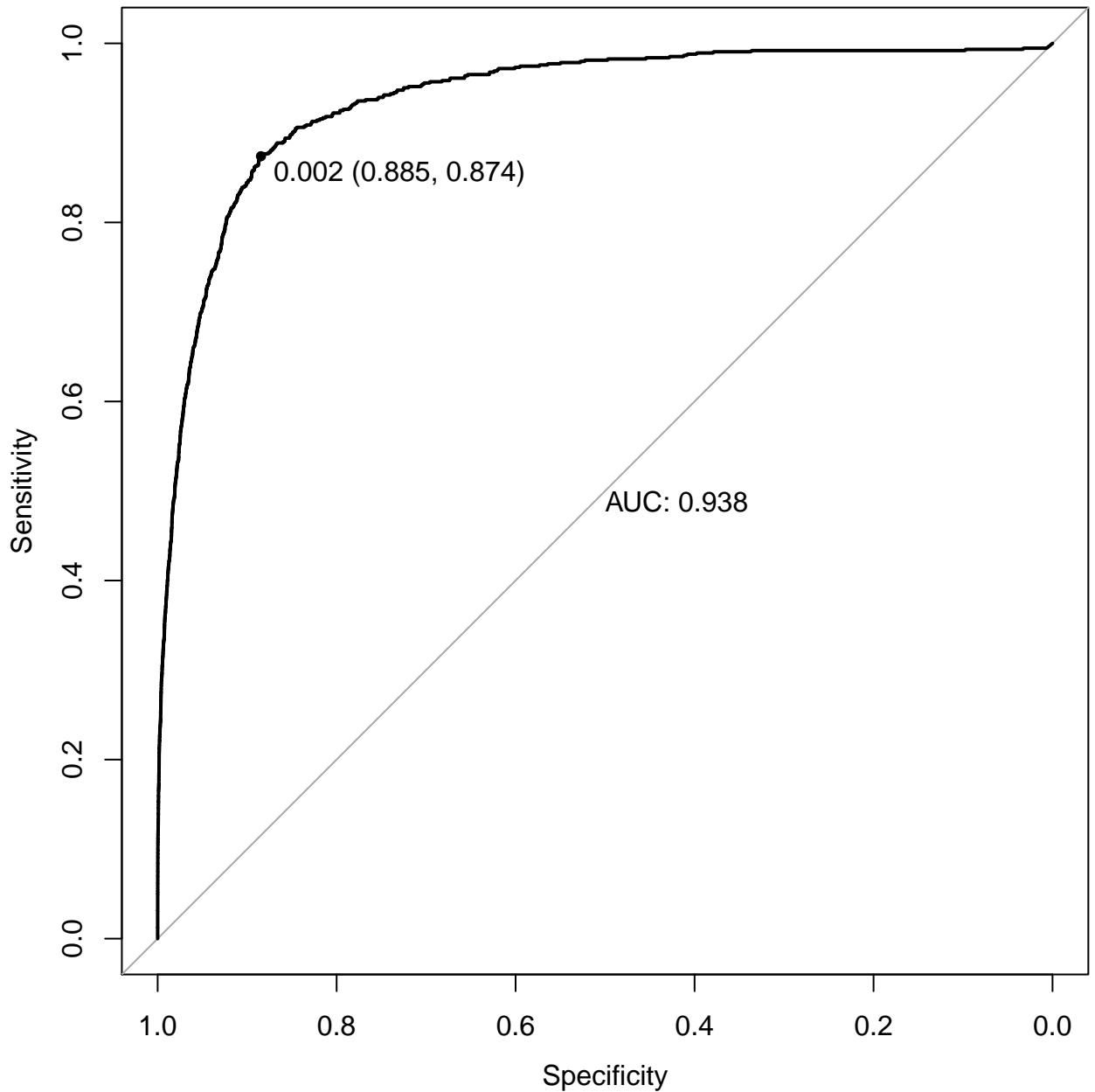
Detection Rate : 0.000483

Detection Prevalence : 0.002624

Balanced Accuracy : 0.608847

'Positive' Class : TRUE

ROC curve for model 12




```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.1122  -0.0474  -0.0441  -0.0428   3.7510

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -7.03418    0.04396 -160.031  < 2e-16 ***
dx_revenue      0.23364    0.02316  10.088  < 2e-16 ***
dx_session_time 0.01385    0.00028  49.443  < 2e-16 ***
dx_gems_count   0.15911    0.02050   7.763 8.31e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 10910  on 501484  degrees of freedom
AIC: 10918

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.9382
```

```
$relativeCountDifference
```

```
[1] 1.151475
```

```
$optimal_cut_off
```

```
[1] 0.06905214
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338123	582
TRUE	695	164

Accuracy : 0.9962

95% CI : (0.996, 0.9964)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 1.000000

Kappa : 0.2025

McNemar's Test P-Value : 0.001723

Sensitivity : 0.219839

Specificity : 0.997949

Pos Pred Value : 0.190920

Neg Pred Value : 0.998282

Prevalence : 0.002197

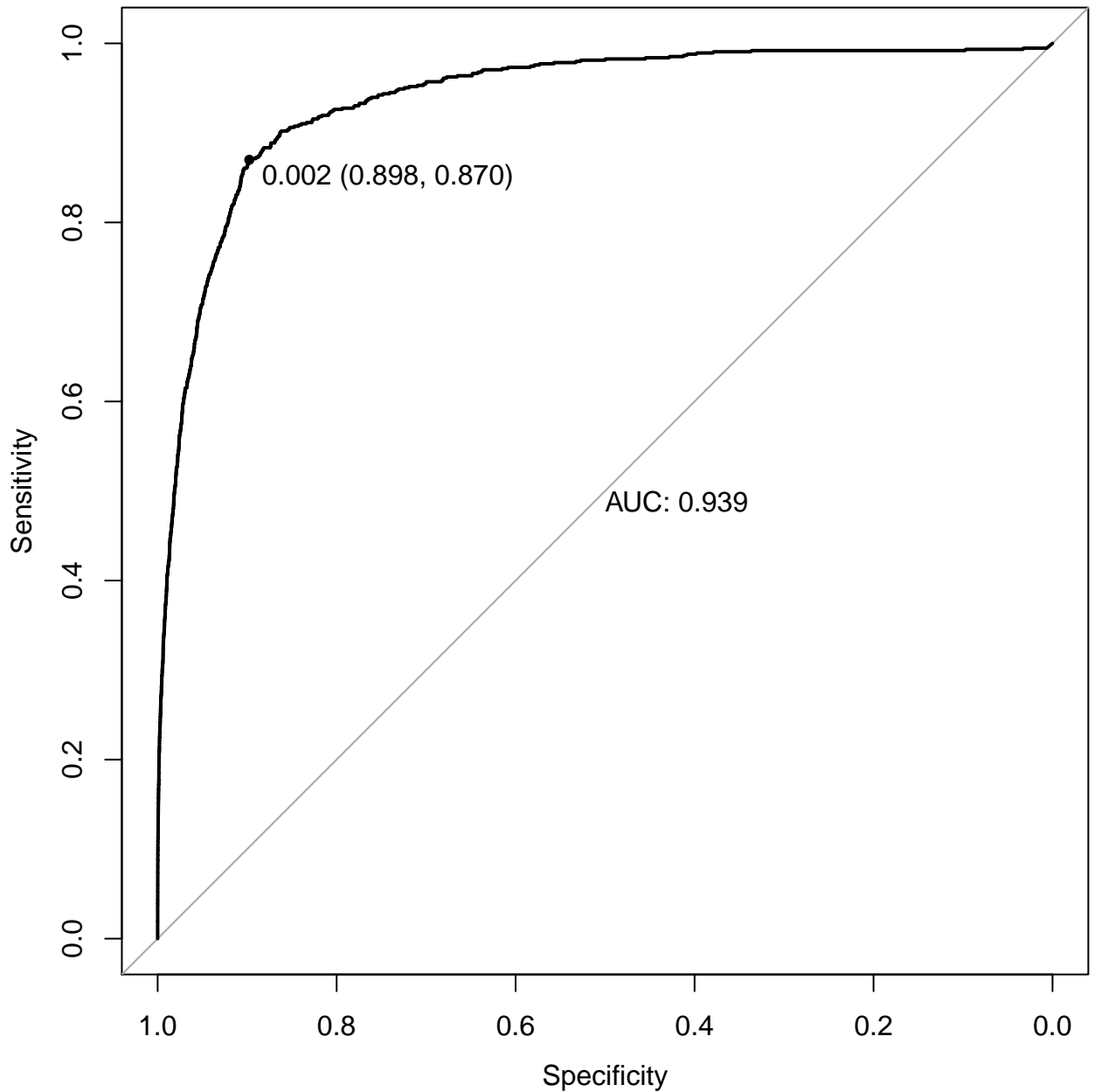
Detection Rate : 0.000483

Detection Prevalence : 0.002530

Balanced Accuracy : 0.608894

'Positive' Class : TRUE

ROC curve for model 13



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.5191  -0.0482  -0.0447  -0.0434   3.7448

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -7.0108389   0.0435062 -161.146  <2e-16 ***
dx_revenue     0.2881713   0.0310145   9.292  <2e-16 ***
dx_session_time 0.0144925   0.0002669  54.290  <2e-16 ***
dx_gems_spent  0.0023429   0.0010802   2.169   0.0301 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669  on 501487  degrees of freedom
Residual deviance: 10964  on 501484  degrees of freedom
AIC: 10972

Number of Fisher Scoring iterations: 9

```

```
$auc
```

```
Area under the curve: 0.9394
```

```
$relativeCountDifference
```

```
[1] 1.190349
```

```
$optimal_cut_off
```

```
[1] 0.06633997
```


\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338097	579
TRUE	721	167

Accuracy : 0.9962
95% CI : (0.996, 0.9964)
No Information Rate : 0.9978
P-Value [Acc > NIR] : 1

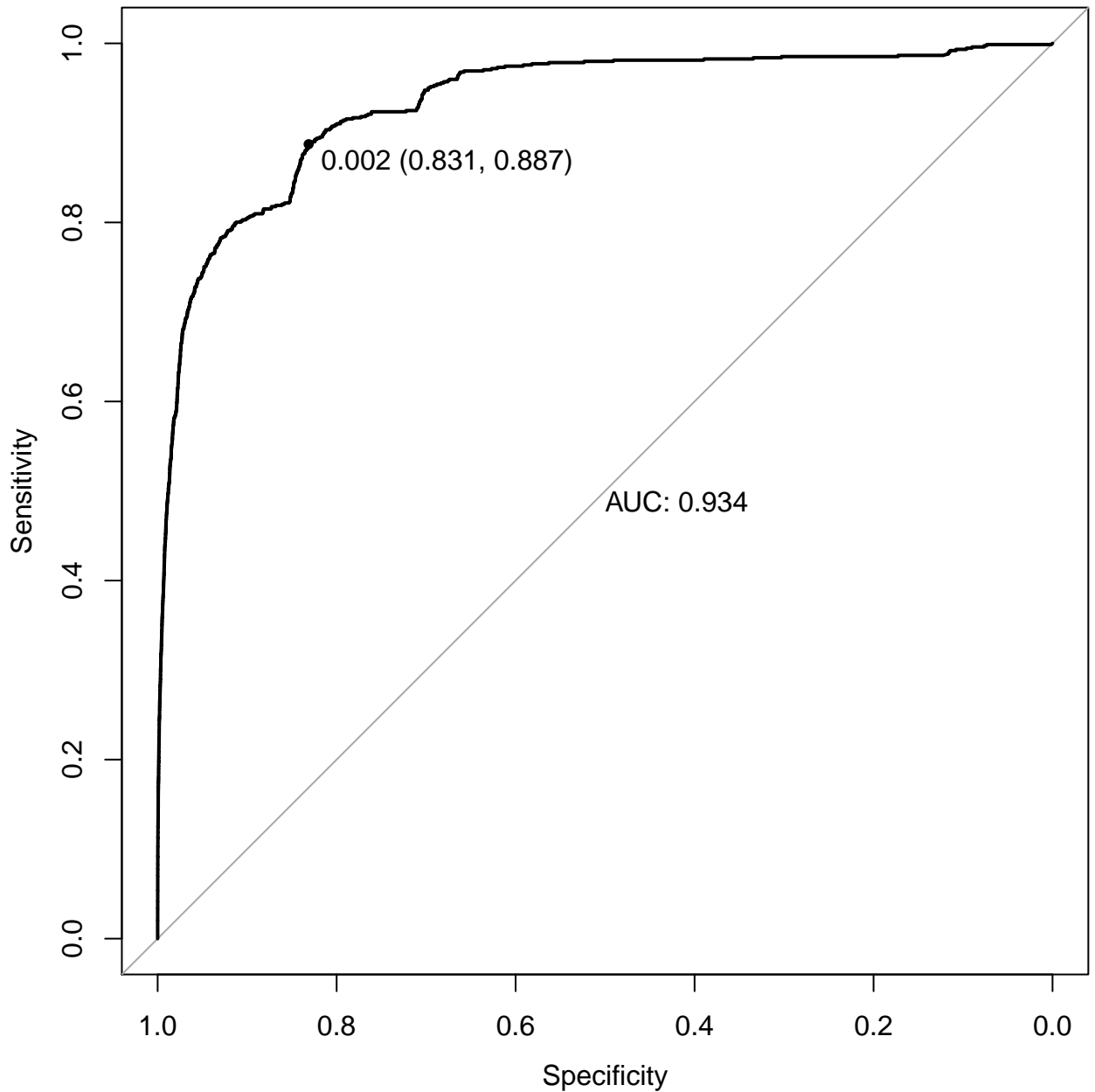
Kappa : 0.2025

McNemar's Test P-Value : 9.205e-05

Sensitivity : 0.2238606
Specificity : 0.9978720
Pos Pred Value : 0.1880631
Neg Pred Value : 0.9982904
Prevalence : 0.0021969
Detection Rate : 0.0004918
Detection Prevalence : 0.0026151
Balanced Accuracy : 0.6108663

'Positive' Class : TRUE

ROC curve for model 14



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.6637  -0.0396  -0.0294  -0.0282   4.1619

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.4361917  0.0791238  -56.07  <2e-16 ***
dx_revenue     0.2516672  0.0181099   13.90  <2e-16 ***
dx_session_time 0.0138757  0.0002716   51.08  <2e-16 ***
tier          -1.1388597  0.0386538  -29.46  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669.0  on 501487  degrees of freedom
Residual deviance:  9972.9  on 501484  degrees of freedom
AIC: 9980.9

Number of Fisher Scoring iterations: 10

```

```
$auc
```

```
Area under the curve: 0.9337
```

```
$relativeCountDifference
```

```
[1] 1.065684
```

```
$optimal_cut_off
```

```
[1] 0.08370598
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338204	565
TRUE	614	181

Accuracy : 0.9965

95% CI : (0.9963, 0.9967)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 1.0000

Kappa : 0.2332

McNemar's Test P-Value : 0.1621

Sensitivity : 0.242627

Specificity : 0.998188

Pos Pred Value : 0.227673

Neg Pred Value : 0.998332

Prevalence : 0.002197

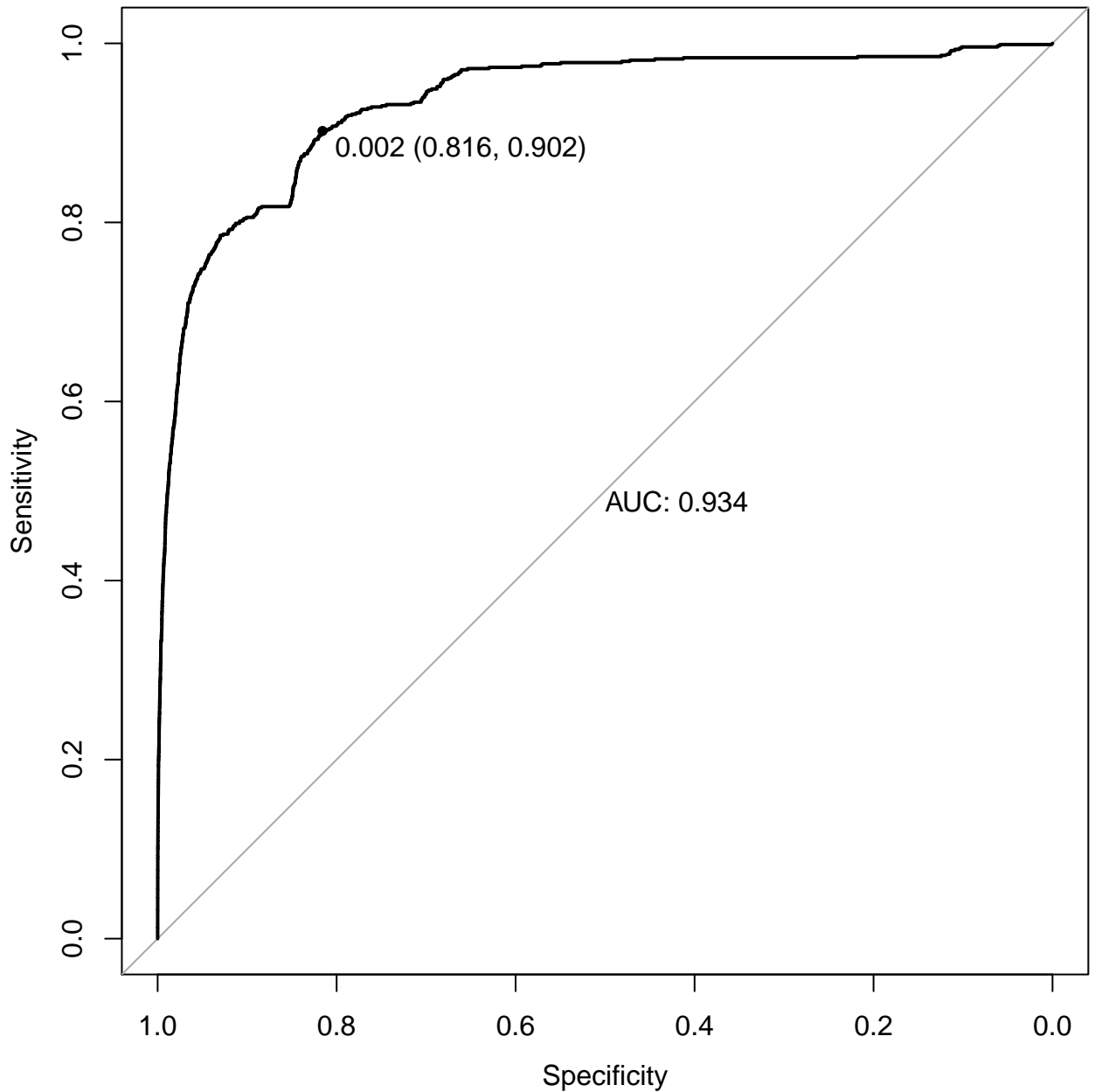
Detection Rate : 0.000533

Detection Prevalence : 0.002341

Balanced Accuracy : 0.620408

'Positive' Class : TRUE

ROC curve for model 15



```

Call:
glm(formula = f, family = "binomial", data = datTrain)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.8824  -0.0402  -0.0286  -0.0273   4.1740

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.585400   0.081223 -56.455 < 2e-16 ***
dx_revenue    -0.033567   0.030502  -1.100  0.27112
dx_pay_count   2.028852   0.182280  11.130 < 2e-16 ***
dx_login_count  0.051520   0.011359   4.536 5.74e-06 ***
dx_session_count 0.054238   0.011747   4.617 3.89e-06 ***
dx_session_time 0.010265   0.000474  21.658 < 2e-16 ***
dx_gems_count   0.160303   0.029680   5.401 6.62e-08 ***
dx_gems_spent  -0.004636   0.001358  -3.414 0.00064 ***
tier          -1.144827   0.039211 -29.196 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 14669.0  on 501487  degrees of freedom
Residual deviance:  9741.3  on 501479  degrees of freedom
AIC: 9759.3

Number of Fisher Scoring iterations: 10

```



```
$auc
```

```
Area under the curve: 0.9344
```

```
$relativeCountDifference
```

```
[1] 1.108579
```

```
$optimal_cut_off
```

```
[1] 0.08830313
```

\$confMatrix

Confusion Matrix and Statistics

	Reference	
Prediction	FALSE	TRUE
FALSE	338172	565
TRUE	646	181

Accuracy : 0.9964

95% CI : (0.9962, 0.9966)

No Information Rate : 0.9978

P-Value [Acc > NIR] : 1.00000

Kappa : 0.2284

McNemar's Test P-Value : 0.02151

Sensitivity : 0.242627

Specificity : 0.998093

Pos Pred Value : 0.218863

Neg Pred Value : 0.998332

Prevalence : 0.002197

Detection Rate : 0.000533

Detection Prevalence : 0.002435

Balanced Accuracy : 0.620360

'Positive' Class : TRUE