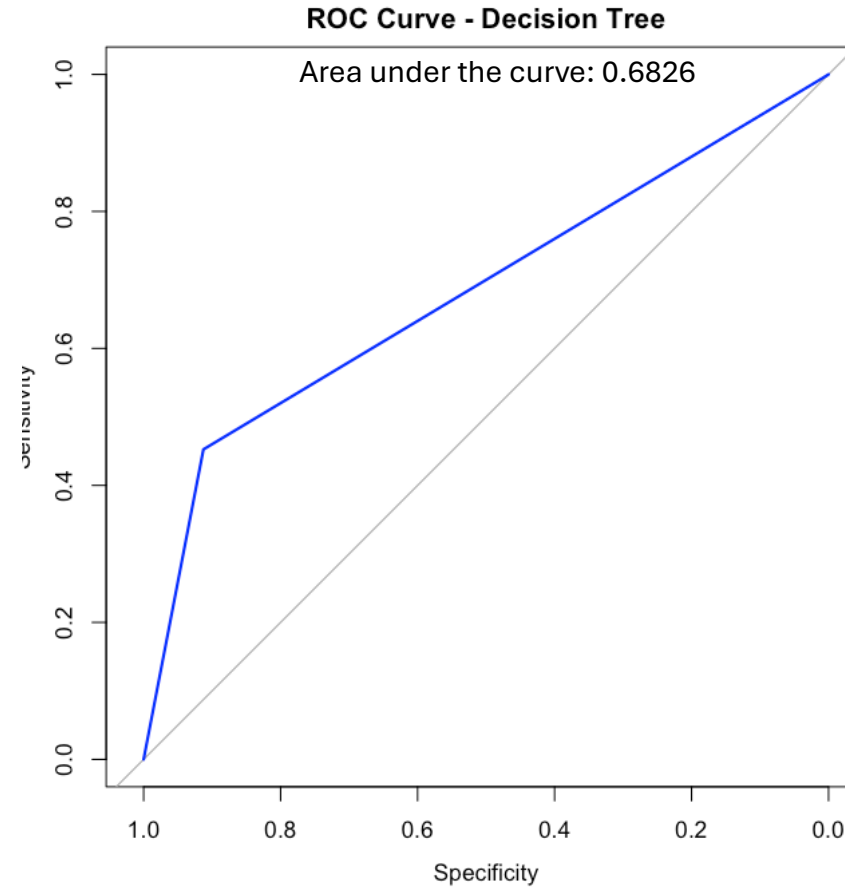# DECISION TREE 1

```
> print(importance)
                        Overall
ap_hi                 1521.4864
ap_lo                 1036.4765
bp_category           1035.8156
bp_category_encoded   1035.8156
cholesterol            311.8565
gender                   0.0000
height                   0.0000
weight                   0.0000
gluc                     0.0000
smoke                    0.0000
alco                     0.0000
active                   0.0000
age_years                0.0000
bmi                      0.0000
```

ROC Curve - Decision Tree

Area under the curve: 0.6826

Specificity

```
Confusion Matrix and Statistics

               Reference
Prediction    no    yes
       no   10989   2168
       yes   1051   1792

               Accuracy : 0.7988
                 95% CI : (0.7925, 0.805)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : < 0.0000000000000022

                  Kappa : 0.4034

 Mcnemar's Test P-Value : < 0.0000000000000022

            Sensitivity : 0.4525
            Specificity : 0.9127
         Pos Pred Value : 0.6303
         Neg Pred Value : 0.8352
             Prevalence : 0.2475
         Detection Rate : 0.1120
   Detection Prevalence : 0.1777
      Balanced Accuracy : 0.6826

       'Positive' Class : yes
```

DECISION TREE TUNED FOR SENSITIVITY

```
> print(importance3)
```

	Overall
age_years	180.295709
ap_hi	1726.823074
ap_lo	1136.611726
bmi	25.055042
bp_category	1098.196674
bp_category_encoded	1038.108592
cholesterol	522.000594
gluc	25.782670
height	4.690020
weight	4.170049
gender	0.000000
smoke	0.000000
alco	0.000000
active	0.000000

ROC Curve - Decision Tree Caret

Area under the curve: 0.7148

(y-axis: Sensitivity, x-axis: Specificity)
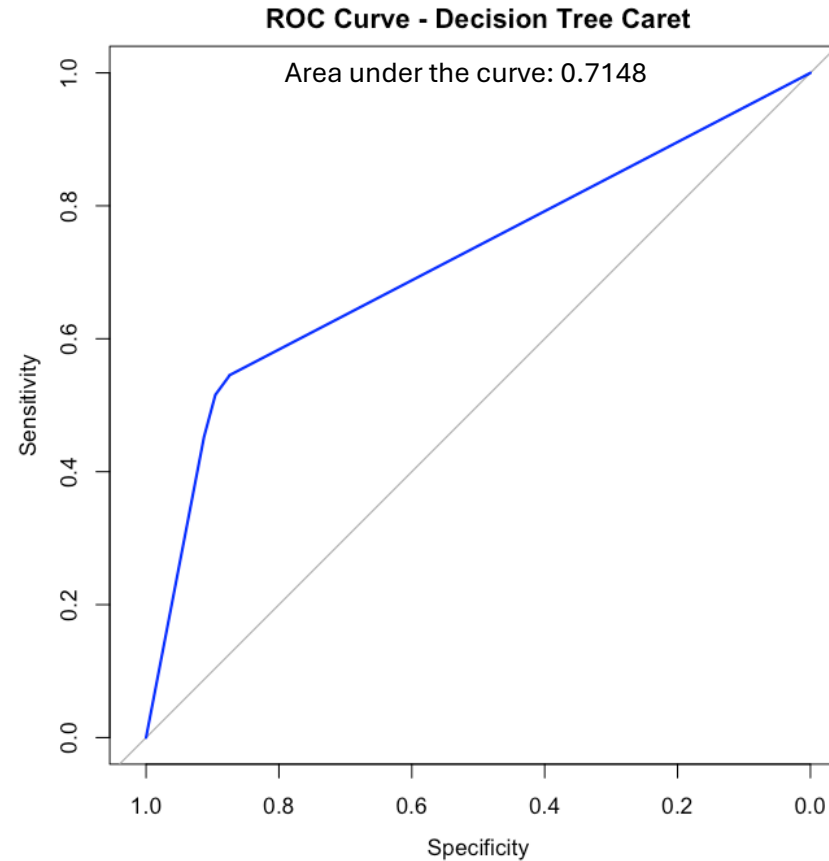
```
                    Reference
Prediction     no     yes
       no    10780    1913
       yes    1260    2047

              Accuracy : 0.8017
                95% CI : (0.7954, 0.8078)
   No Information Rate : 0.7525
   P-Value [Acc > NIR] : < 0.00000000000000022

                 Kappa : 0.4364

 Mcnemar's Test P-Value : < 0.00000000000000022

           Sensitivity : 0.5169
           Specificity : 0.8953
        Pos Pred Value : 0.6190
        Neg Pred Value : 0.8493
            Prevalence : 0.2475
        Detection Rate : 0.1279
  Detection Prevalence : 0.2067
     Balanced Accuracy : 0.7061

      'Positive' Class : yes
```
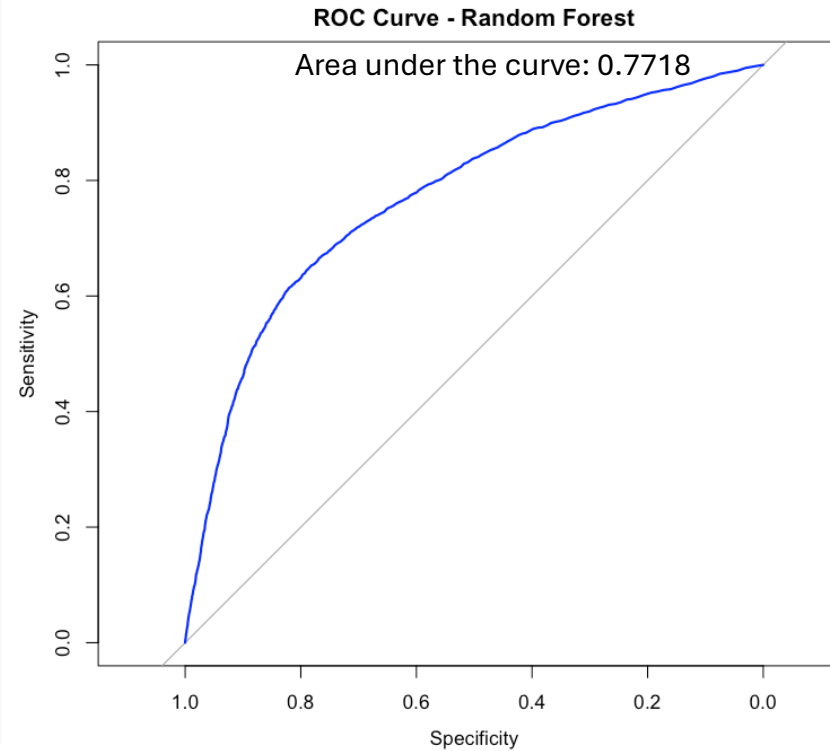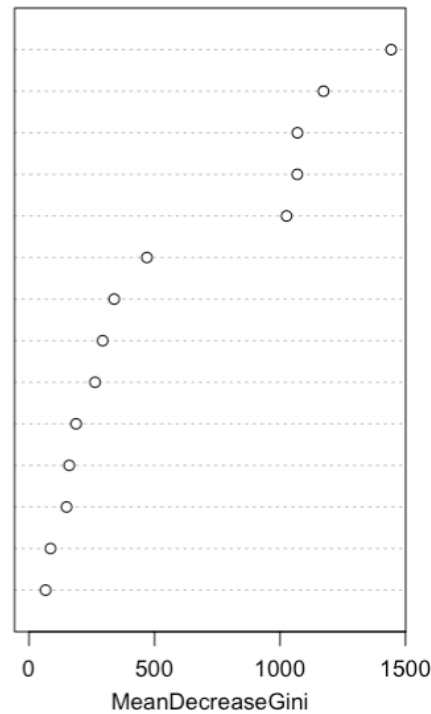
RANDOM FOREST (ntree = 500; No. of variables tried at each split: 4)

| bmi |
| ap_hi |
| age_years |
| weight |
| height |
| ap_lo |
| cholesterol |
| bp_category_encoded |
| bp_category |
| gluc |
| gender |
| active |
| smoke |
| alco |

MeanDecreaseGini

ROC Curve - Random Forest

Area under the curve: 0.7718

Sensitivity / Specificity

```
Confusion Matrix and Statistics

                Reference
Prediction     no    yes
       no   10914   2191
       yes   1126   1769

               Accuracy : 0.7927
                 95% CI : (0.7863, 0.7989)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.3882

 Mcnemar's Test P-Value : < 0.00000000000000022

            Sensitivity : 0.4467
            Specificity : 0.9065
         Pos Pred Value : 0.6111
         Neg Pred Value : 0.8328
             Prevalence : 0.2475
         Detection Rate : 0.1106
   Detection Prevalence : 0.1809
      Balanced Accuracy : 0.6766

       'Positive' Class : yes
```

NAIVE BAYES (normale e kernel density estimation)

```
> print(conf_matrix_norm)
Confusion Matrix and Statistics

          Reference
Prediction    1     0
         1  1928  1504
         0  2032 10536

               Accuracy : 0.779
                 95% CI : (0.7725, 0.7854)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : 2.003e-15

                  Kappa : 0.3789

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.4869
            Specificity : 0.8751
         Pos Pred Value : 0.5618
         Neg Pred Value : 0.8383
             Prevalence : 0.2475
         Detection Rate : 0.1205
   Detection Prevalence : 0.2145
      Balanced Accuracy : 0.6810

       'Positive' Class : 1
```

```
> print(conf_matrix_kernel)
Confusion Matrix and Statistics

          Reference
Prediction    1     0
         1  1732  1303
         0  2228 10737

               Accuracy : 0.7793
                 95% CI : (0.7728, 0.7857)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : 9.419e-16

                  Kappa : 0.3571

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.4374
            Specificity : 0.8918
         Pos Pred Value : 0.5707
         Neg Pred Value : 0.8282
             Prevalence : 0.2475
         Detection Rate : 0.1082
   Detection Prevalence : 0.1897
      Balanced Accuracy : 0.6646

       'Positive' Class : 1
```
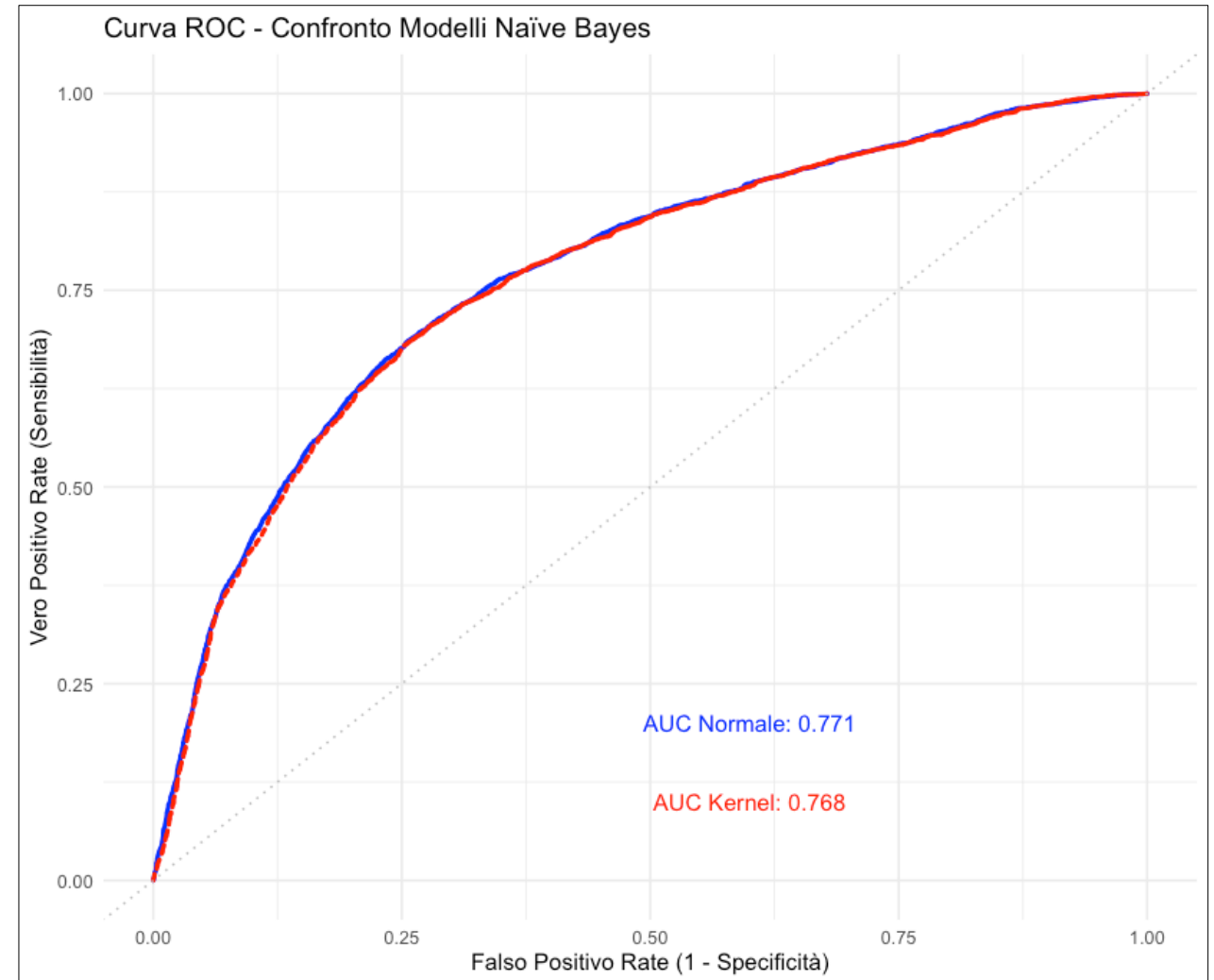
Curva ROC - Confronto Modelli Naïve Bayes

AUC Normale: 0.771

AUC Kernel: 0.768

QDA

```
Single term deletions

Model:
cardio ~ weight + ap_hi + ap_lo + cholesterol + gluc + smoke +
    alco + active + age_years + bp_category
            Df Deviance  F value      Pr(>F)
<none>         22278
weight       1   22317   42.0540  9.052e-11 ***
ap_hi        1   23038  818.3020  < 2.2e-16 ***
ap_lo        1   22282    4.1071  0.0427141 *
cholesterol  1   22522  262.9126  < 2.2e-16 ***
gluc         1   22290   13.3802  0.0002548 ***
smoke        1   22287   10.2024  0.0014044 **
alco         1   22290   13.4441  0.0002463 ***
active       1   22306   30.6818  3.072e-08 ***
age_years    1   22604  350.7295  < 2.2e-16 ***
bp_category  3   22343   23.3584  4.323e-15 ***
```

Curva ROC - AUC: 0.69
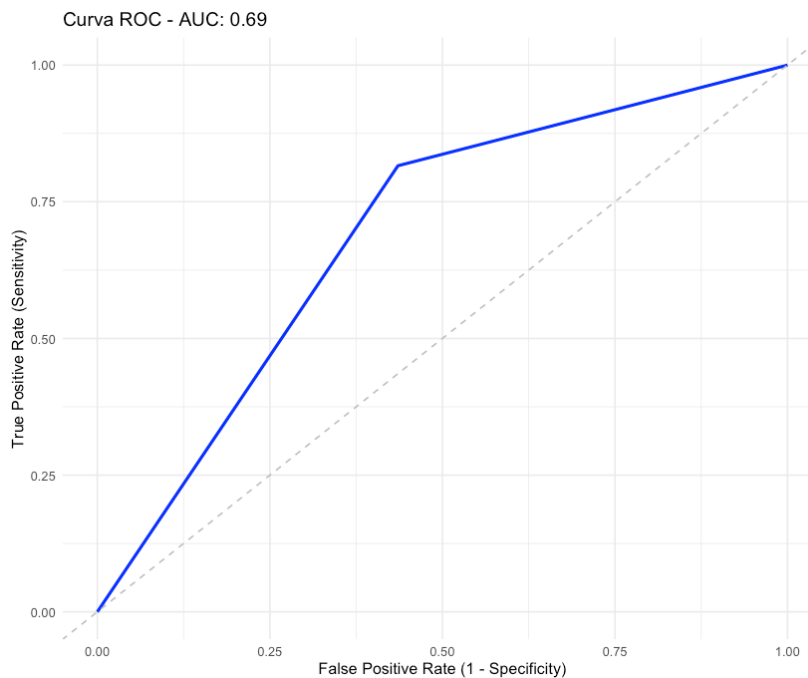

```
Confusion Matrix and Statistics

           Reference
Prediction    1    0
         1 2235 2217
         0 1725 9823

               Accuracy : 0.7536
                 95% CI : (0.7469, 0.7603)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : 0.3748

                  Kappa : 0.365

 Mcnemar's Test P-Value : 5.27e-15

            Sensitivity : 0.5644
            Specificity : 0.8159
         Pos Pred Value : 0.5020
         Neg Pred Value : 0.8506
             Prevalence : 0.2475
         Detection Rate : 0.1397
   Detection Prevalence : 0.2782
      Balanced Accuracy : 0.6901

       'Positive' Class : 1
```
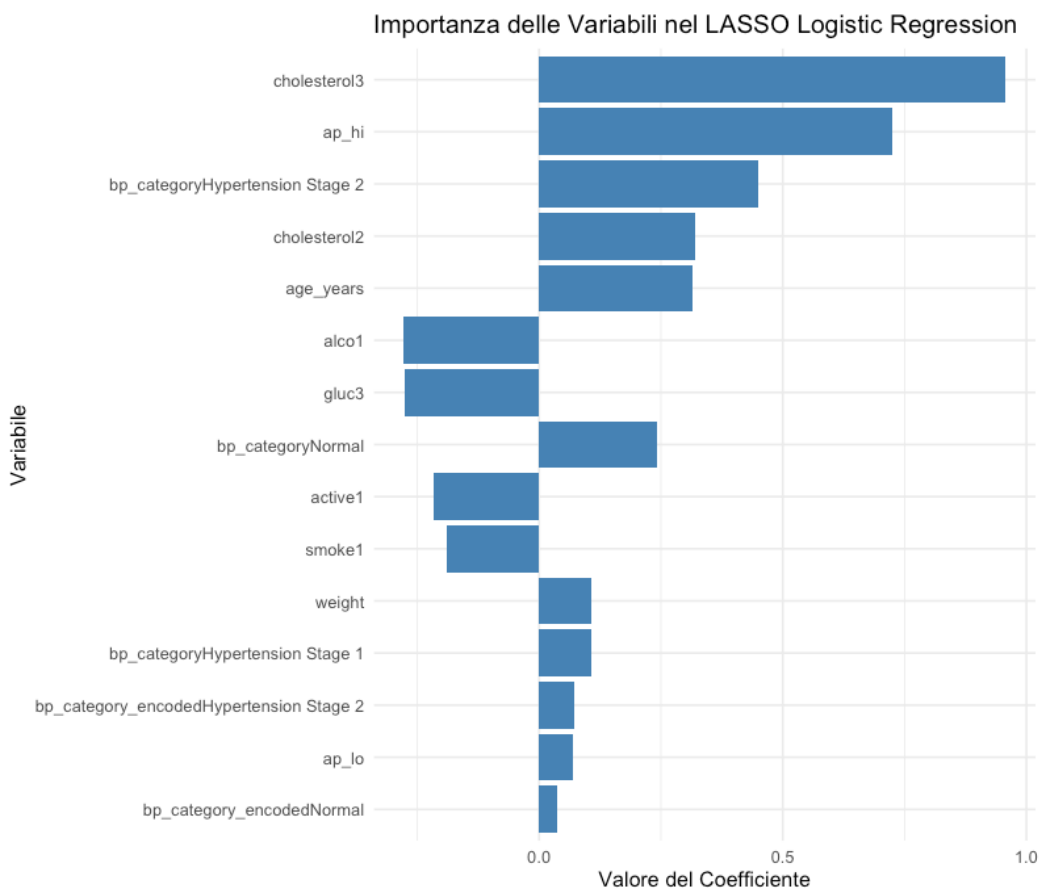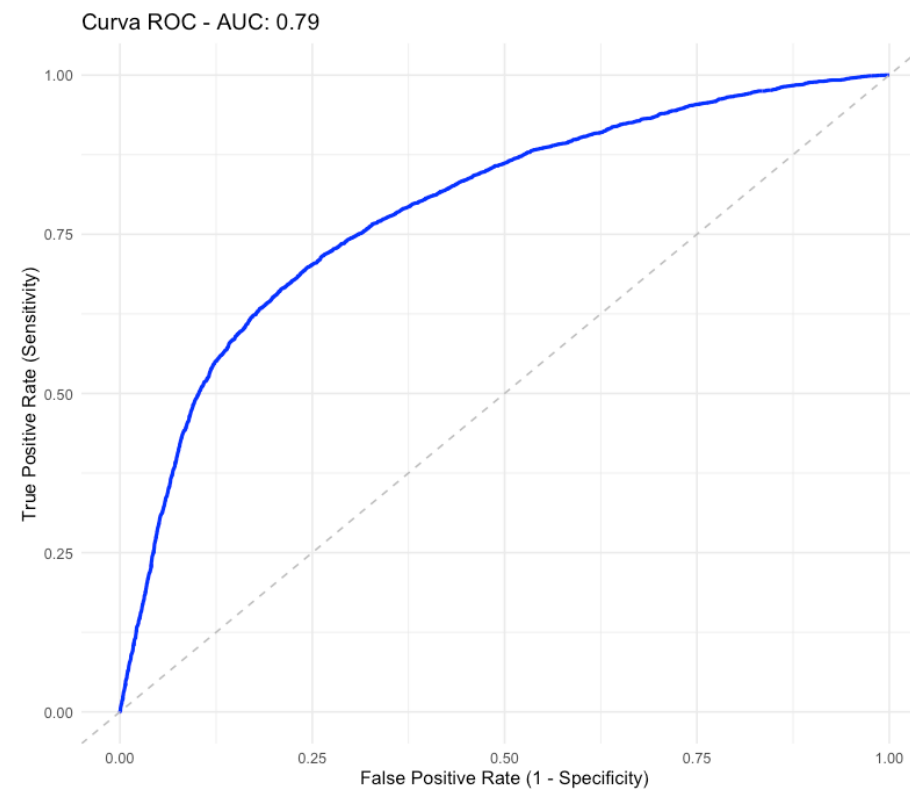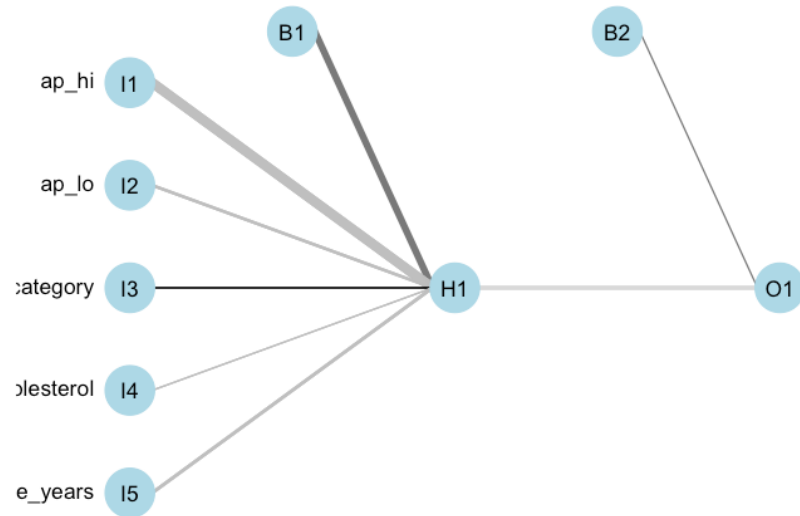
LOGISTIC LASSO (Miglior valore di lambda: 0.0003789426)

Importanza delle Variabili nel LASSO Logistic Regression

Accuracy: 0.7919
Precision: 0.646
Sensitivity: 0.352
F1-score: 0.4557

Curva ROC - AUC: 0.79

PERCEPTRON


```
Confusion Matrix and Statistics

              Reference
Prediction     0     1
         0 10949  2165
         1  1091  1795

               Accuracy : 0.7965
                 95% CI : (0.7902, 0.8027)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : < 0.0000000000000022

                  Kappa : 0.399

 Mcnemar's Test P-Value : < 0.0000000000000022

            Sensitivity : 0.4533
            Specificity : 0.9094
         Pos Pred Value : 0.6220
         Neg Pred Value : 0.8349
             Prevalence : 0.2475
         Detection Rate : 0.1122
   Detection Prevalence : 0.1804
      Balanced Accuracy : 0.6813

       'Positive' Class : 1
```
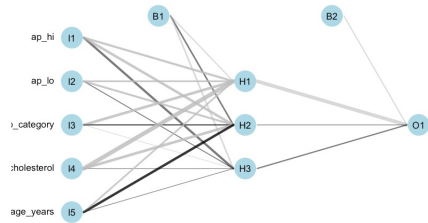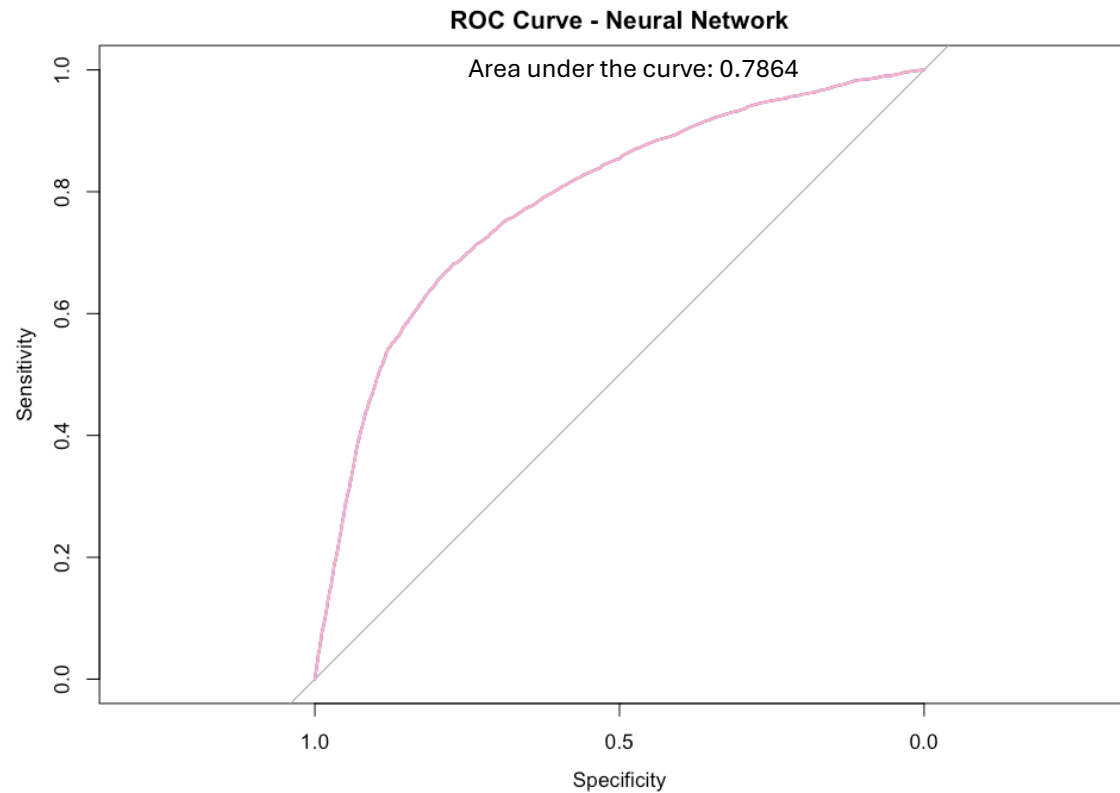
RETE NEURALE MLP


```
> print(importance)
     Overall
X1 34.96618
X2 16.03156
X3 11.32777
X4 23.11674
X5 14.55775
```

ROC Curve - Neural Network

Area under the curve: 0.7864


```
Confusion Matrix and Statistics

          Reference
Prediction     0      1
         0 10984   2177
         1  1056   1783

               Accuracy : 0.7979
                 95% CI : (0.7916, 0.8041)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.4006

 Mcnemar's Test P-Value : < 0.00000000000000022

            Sensitivity : 0.4503
            Specificity : 0.9123
         Pos Pred Value : 0.6280
         Neg Pred Value : 0.8346
             Prevalence : 0.2475
         Detection Rate : 0.1114
   Detection Prevalence : 0.1774
      Balanced Accuracy : 0.6813

       'Positive' Class : 1
```
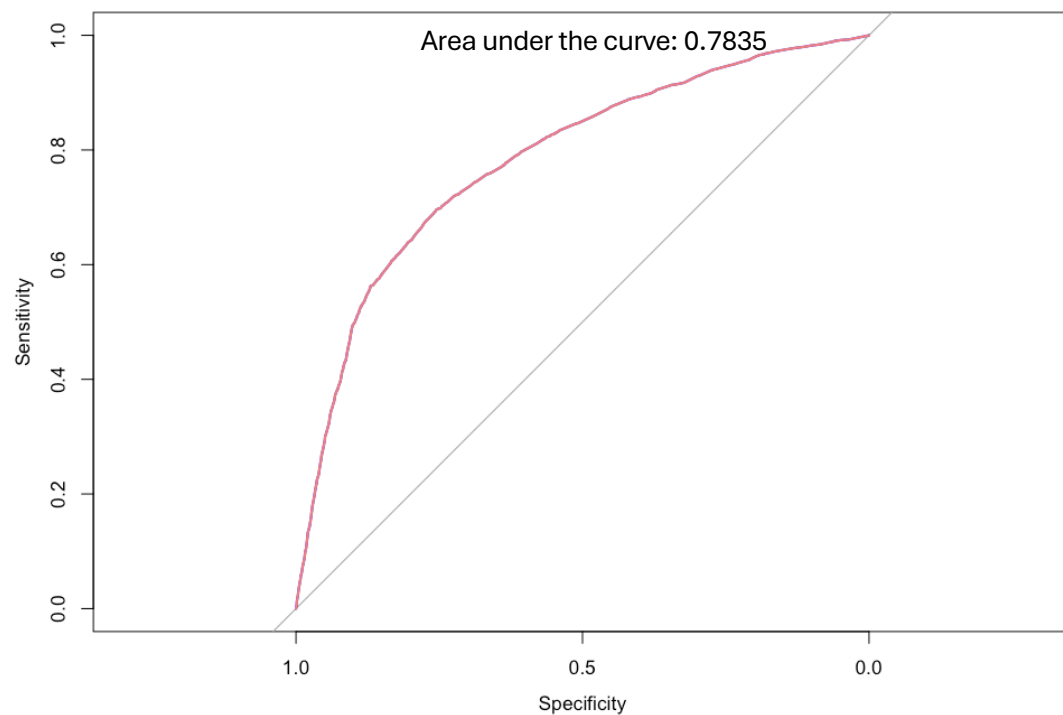
KNN

ROC Curve - KNN

Area under the curve: 0.7835


```
Confusion Matrix and Statistics

          Reference
Prediction    no    yes
       no   10901   2087
       yes   1139   1873

               Accuracy : 0.7984
                 95% CI : (0.7921, 0.8046)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.4114

 Mcnemar's Test P-Value : < 0.00000000000000022

            Sensitivity : 0.4730
            Specificity : 0.9054
         Pos Pred Value : 0.6218
         Neg Pred Value : 0.8393
             Prevalence : 0.2475
         Detection Rate : 0.1171
   Detection Prevalence : 0.1883
      Balanced Accuracy : 0.6892

       'Positive' Class : yes
```
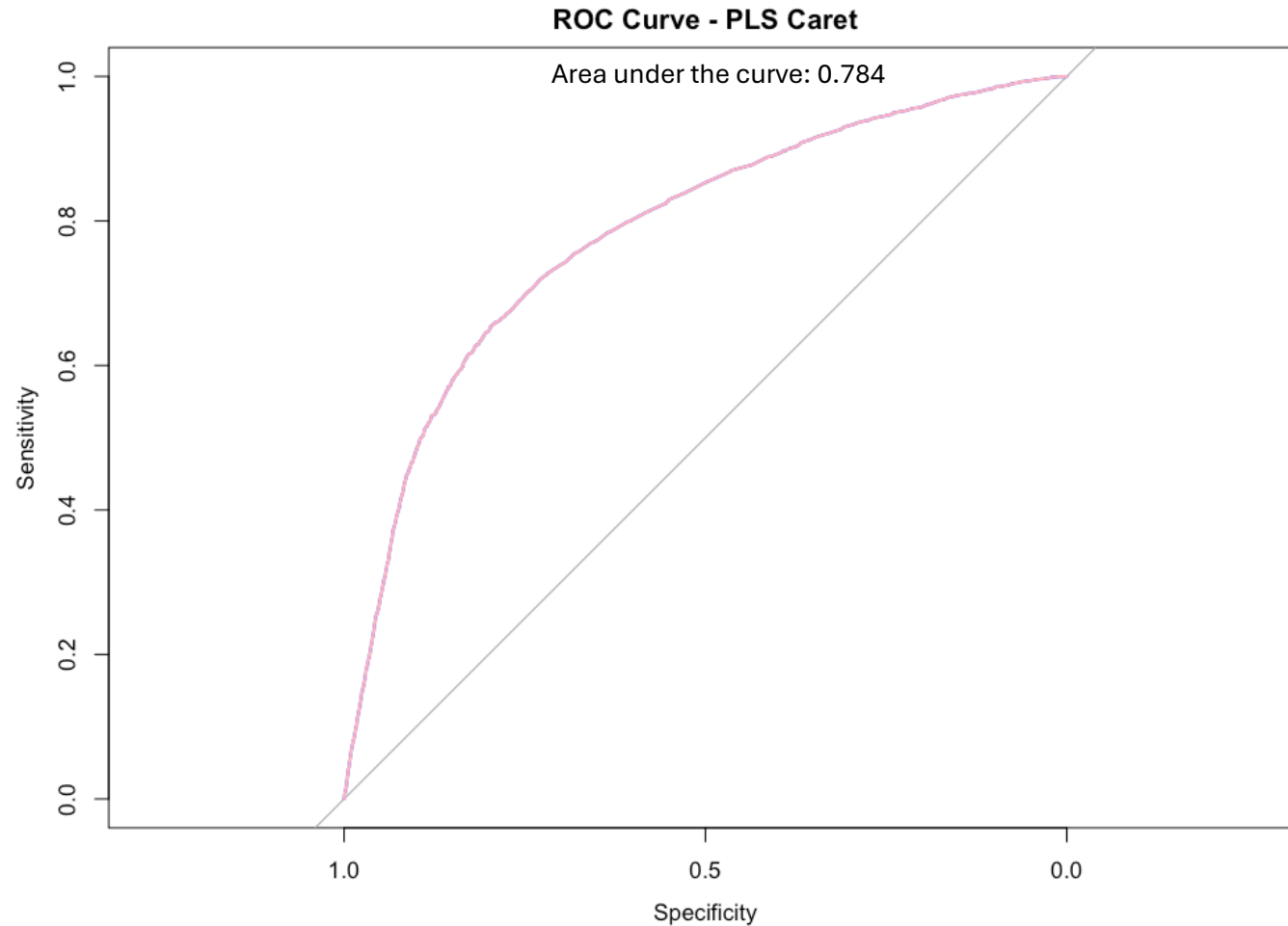
PLS

ROC Curve - PLS Caret

Area under the curve: 0.784


```
Confusion Matrix and Statistics

            Reference
Prediction     1       0
         1   1145     632
         0   2815   11408

               Accuracy : 0.7846
                 95% CI : (0.7781, 0.7909)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.2904

 Mcnemar's Test P-Value : < 0.00000000000000022

            Sensitivity : 0.28914
            Specificity : 0.94751
         Pos Pred Value : 0.64434
         Neg Pred Value : 0.80208
             Prevalence : 0.24750
         Detection Rate : 0.07156
   Detection Prevalence : 0.11106
      Balanced Accuracy : 0.61832

       'Positive' Class : 1
```
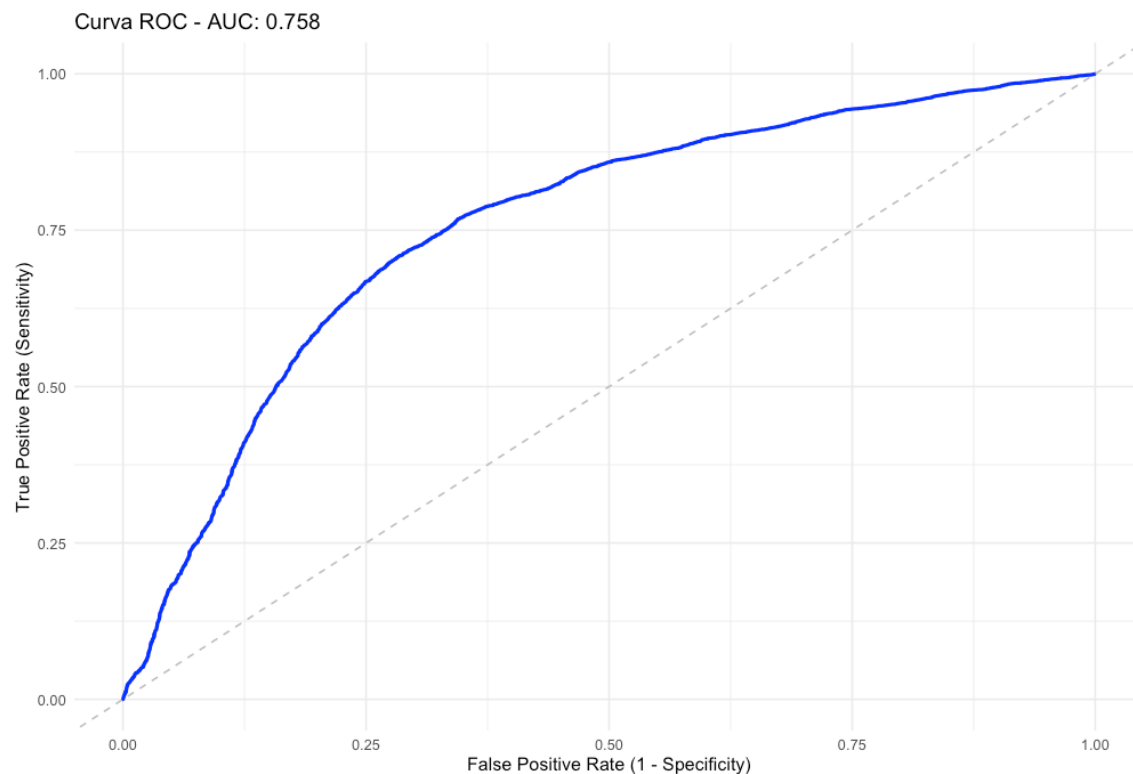
MIGLIOR MODELLO PER SENSITIVITY: QDA

I RISULTATI SEGUENTI SEGUONO CROSS VALIDATION PER MIGLIOR SOGLIA PER MASSIMIZZARE F1_SCORE

"Miglior soglia per massimizzare F1-score: 0.1"

Curva ROC - AUC: 0.758

```
Confusion Matrix and Statistics

            Reference
Prediction     1      0
         1  2295  10663
         0  1665   1377

              Accuracy : 0.2295
                95% CI : (0.223, 0.2361)
    No Information Rate : 0.7525
    P-Value [Acc > NIR] : 1

                 Kappa : -0.1737

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

           Sensitivity : 0.5795
           Specificity : 0.1144
        Pos Pred Value : 0.1771
        Neg Pred Value : 0.4527
            Prevalence : 0.2475
        Detection Rate : 0.1434
  Detection Prevalence : 0.8099
     Balanced Accuracy : 0.3470

      'Positive' Class : 1
```
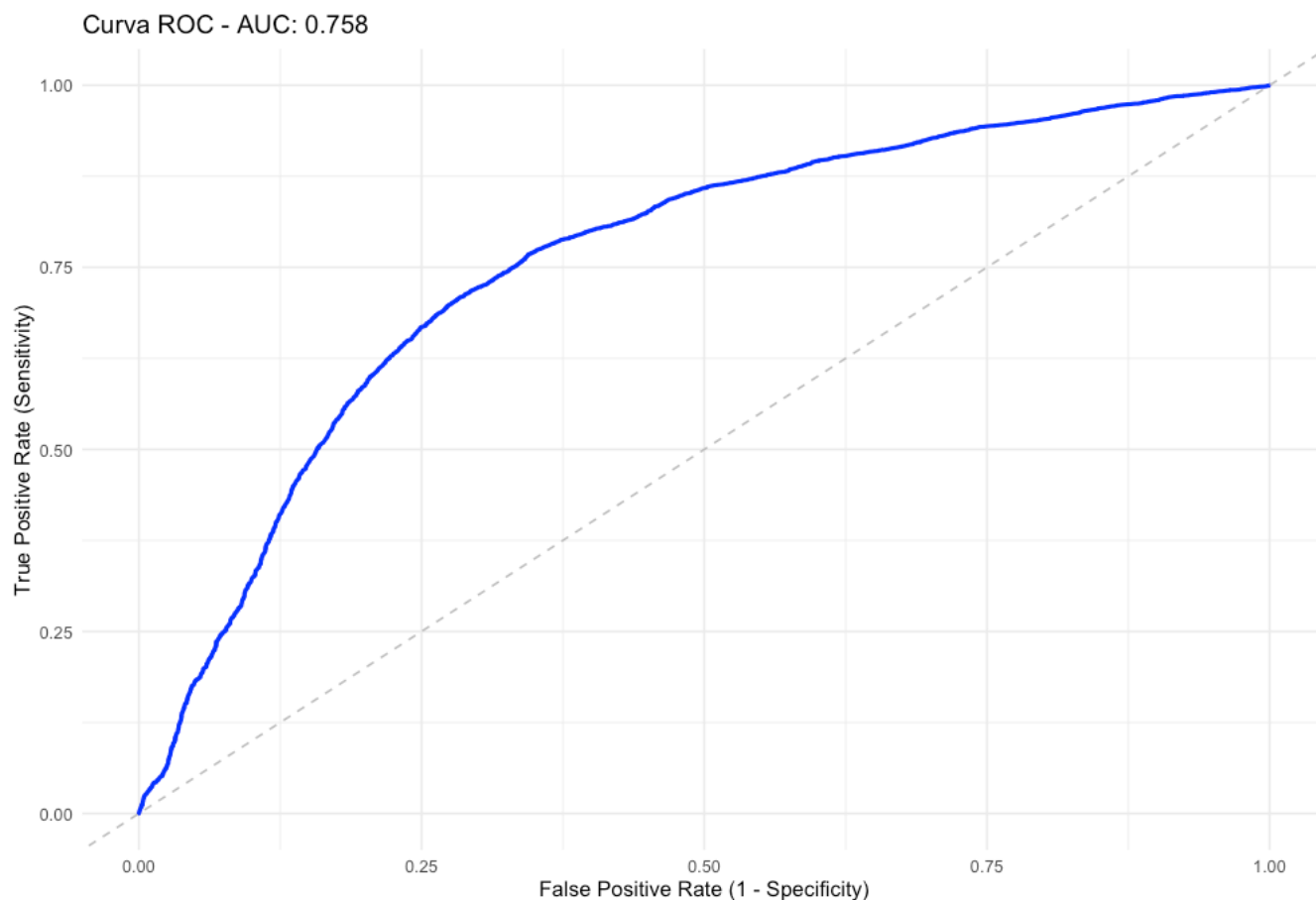
MIGLIOR MODELLO PER SENSITIVITY: QDA

I RISULTATI SEGUENTI SEGUONO CROSS VALIDATION PER MIGLIOR SOGLIA PER MASSIMIZZARE Area Under the Curve: TP/FN

"Miglior soglia per massimizzare AUC: 0.75"

Curva ROC - AUC: 0.758

```
Confusion Matrix and Statistics

                   Reference
Prediction    1      0
         1 1358  9207
         0 2602  2833

              Accuracy : 0.2619
                95% CI : (0.2551, 0.2688)
   No Information Rate : 0.7525
   P-Value [Acc > NIR] : 1

                 Kappa : -0.2704

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

           Sensitivity : 0.34293
           Specificity : 0.23530
        Pos Pred Value : 0.12854
        Neg Pred Value : 0.52125
            Prevalence : 0.24750
        Detection Rate : 0.08488
  Detection Prevalence : 0.66031
     Balanced Accuracy : 0.28911

      'Positive' Class : 1
```