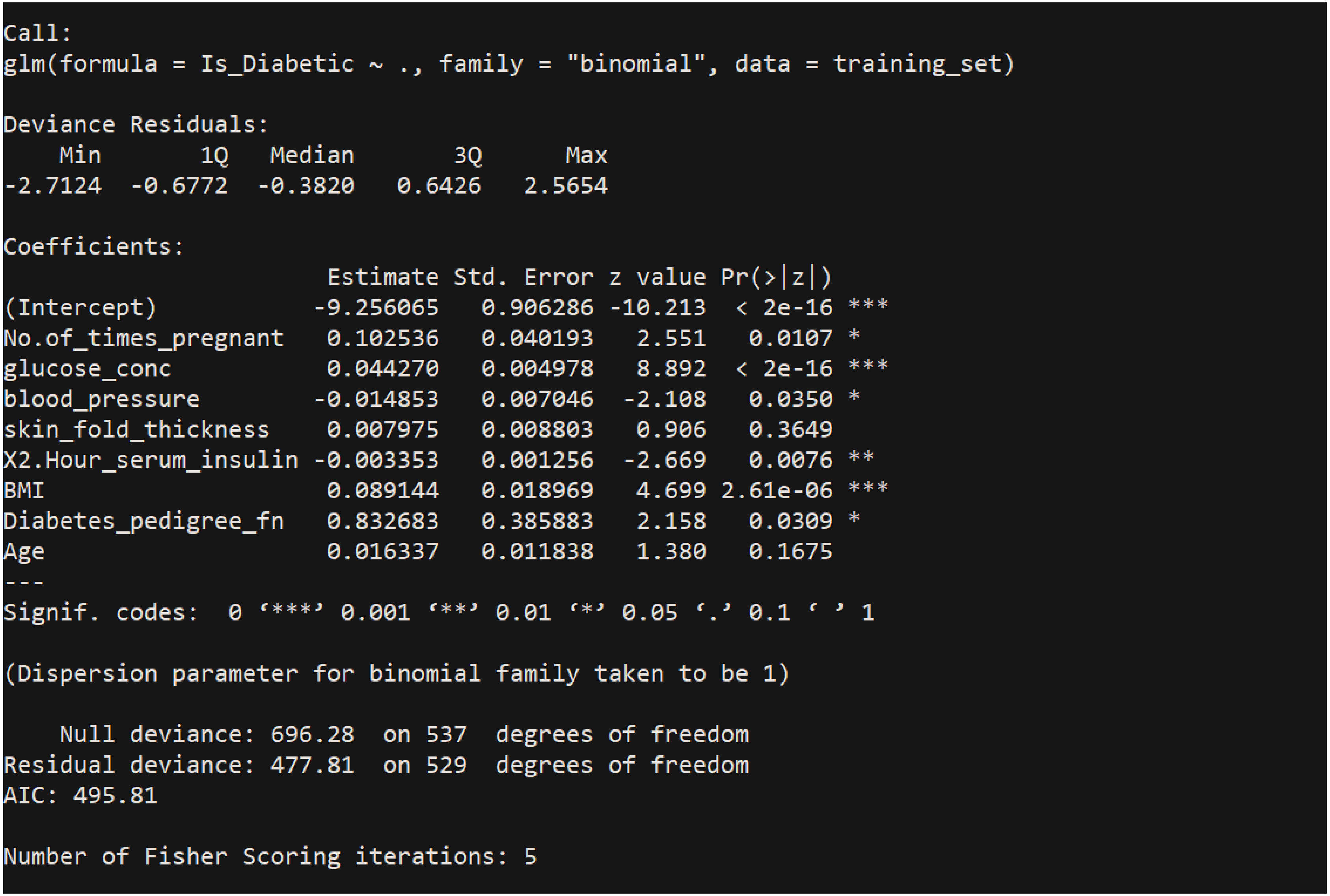
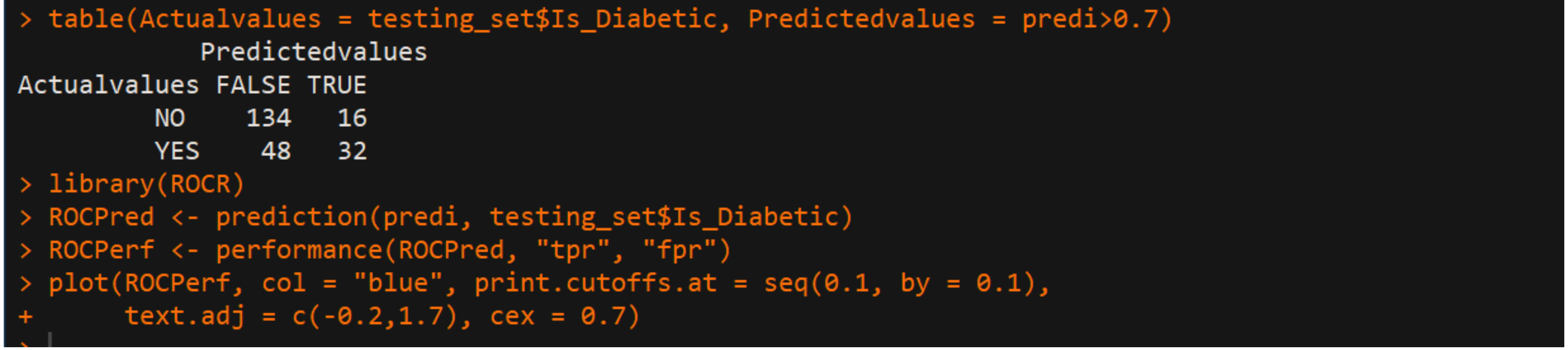
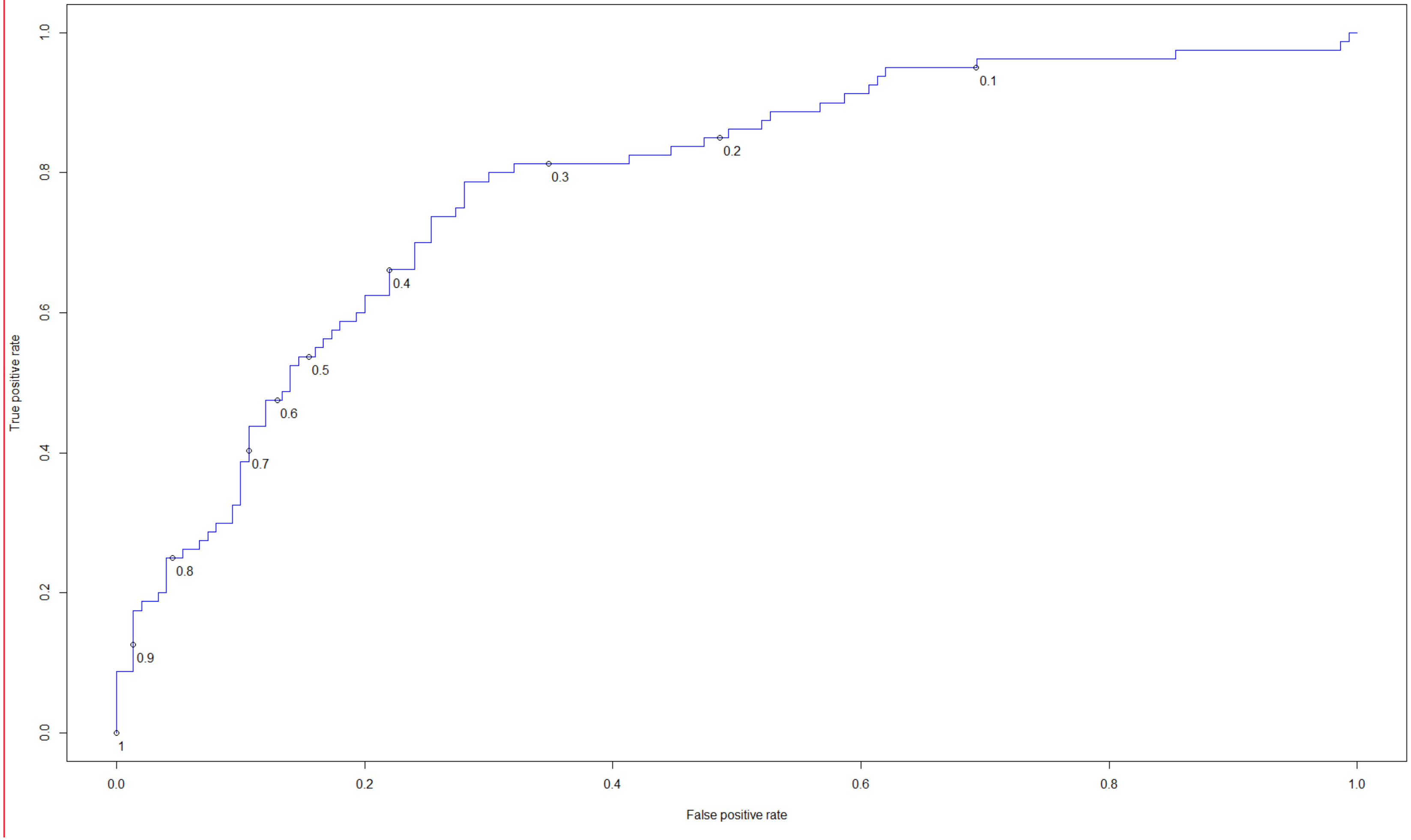
Logistic Regression:



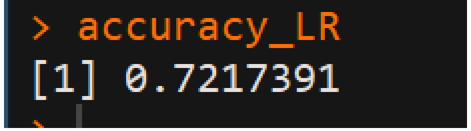
```
Call:
glm(formula = Is_Diabetic ~ ., family = "binomial", data = training_set)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-2.7124  -0.6772  -0.3820   0.6426   2.5654  

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)    
(Intercept)           -9.256065   0.906286 -10.213  < 2e-16 ***
No.of_times_pregnant   0.102536   0.040193   2.551   0.0107 *  
glucose_conc           0.044270   0.004978   8.892  < 2e-16 ***
blood_pressure        -0.014853   0.007046  -2.108   0.0350 *  
skin_fold_thickness    0.007975   0.008803   0.906   0.3649    
X2.Hour_serum_insulin -0.003353   0.001256  -2.669   0.0076 ** 
BMI                    0.089144   0.018969   4.699 2.61e-06 ***
Diabetes_pedigree_fn   0.832683   0.385883   2.158   0.0309 *  
Age                    0.016337   0.011838   1.380   0.1675    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 696.28  on 537  degrees of freedom
Residual deviance: 477.81  on 529  degrees of freedom
AIC: 495.81

Number of Fisher Scoring iterations: 5
```



```
> table(Actualvalues = testing_set$Is_Diabetic, Predictedvalues = predi>0.7)
            Predictedvalues
Actualvalues FALSE TRUE
         NO    134   16
         YES    48   32
> library(ROCR)
> ROCPred <- prediction(predi, testing_set$Is_Diabetic)
> ROCPerf <- performance(ROCPred, "tpr", "fpr")
> plot(ROCPerf, col = "blue", print.cutoffs.at = seq(0.1, by = 0.1),
+      text.adj = c(-0.2,1.7), cex = 0.7)
>
```

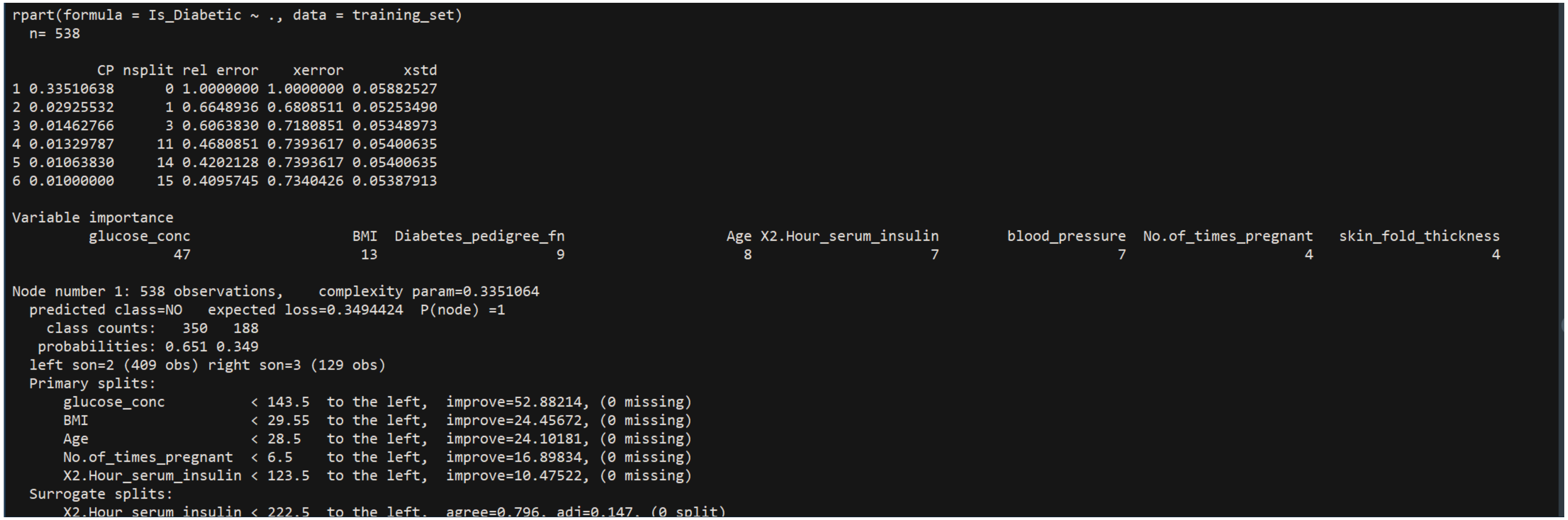


True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

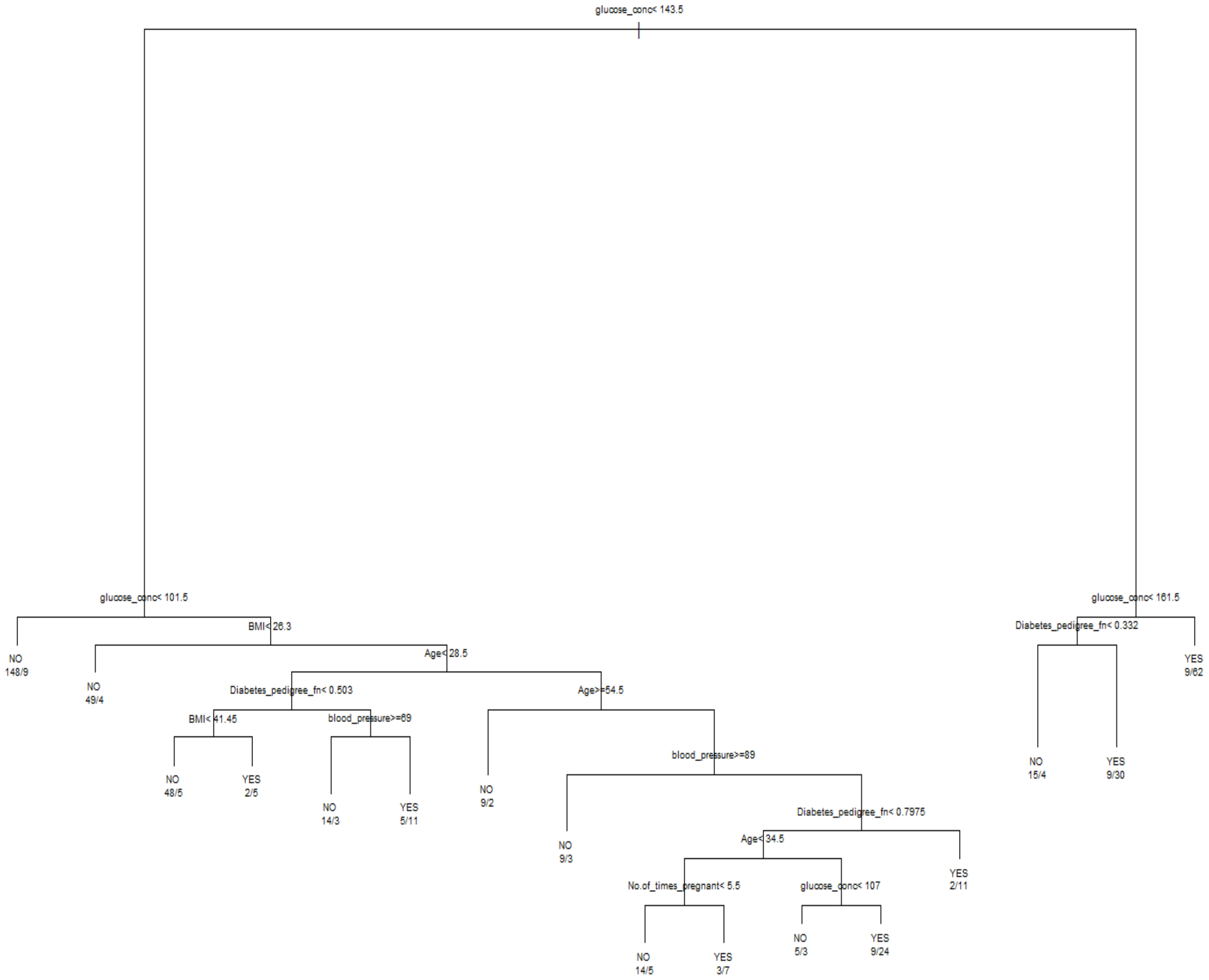


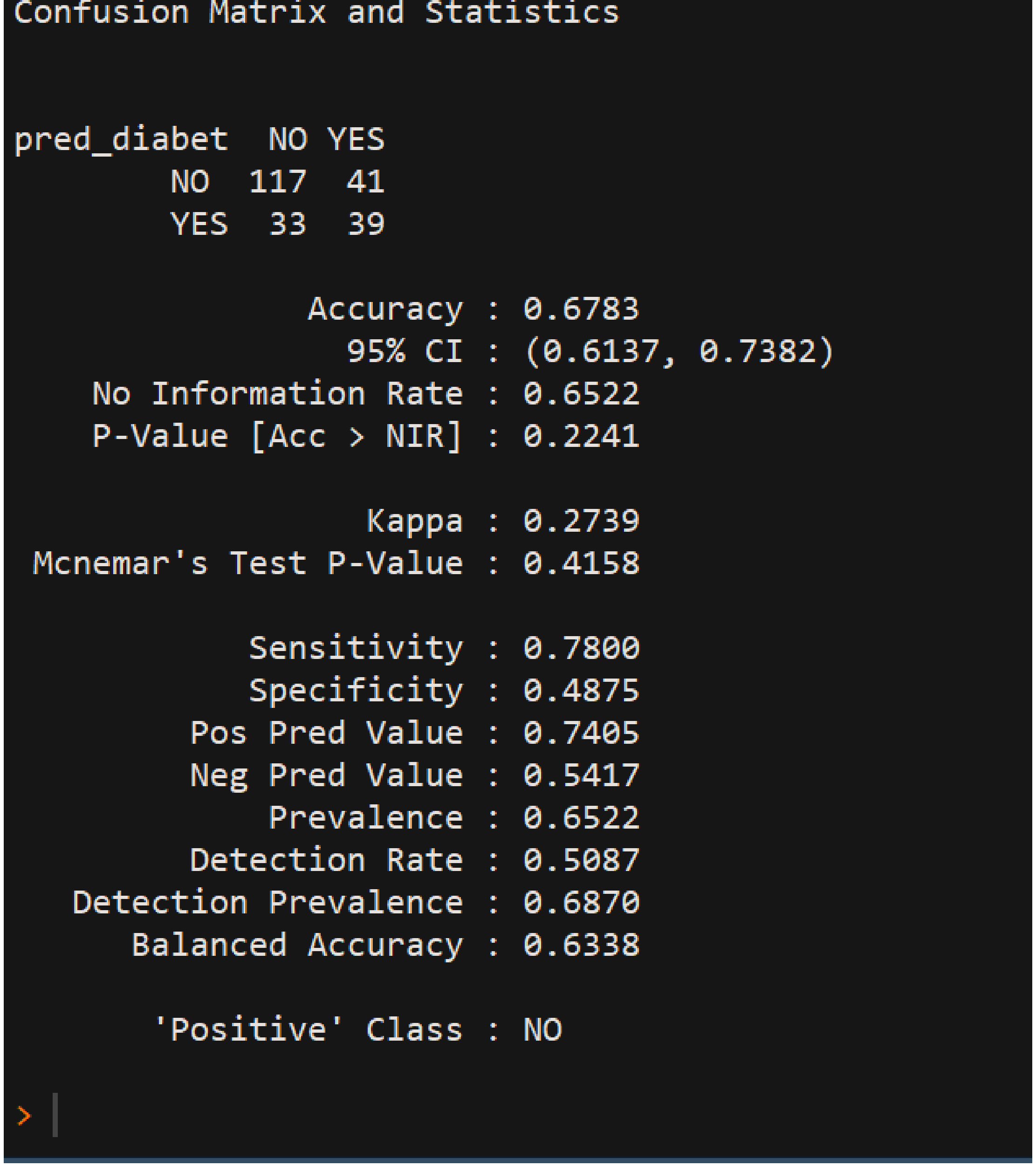
```
> accuracy_LR
[1] 0.7217391
>
```

Decision Tree:



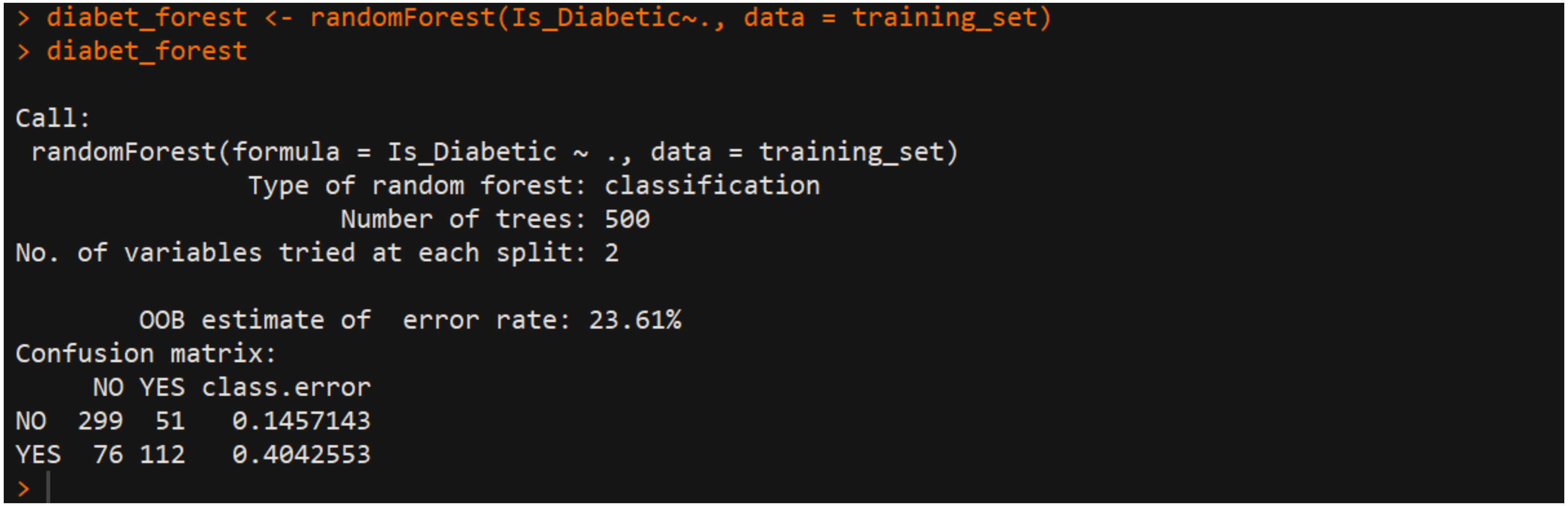
```
rpart(formula = Is_Diabetic ~ ., data = training_set)
  n= 538 

          CP nsplit rel error    xerror       xstd
1 0.33510638      0 1.0000000 1.0000000 0.05882527
2 0.02925532      1 0.6648936 0.6808511 0.05253490
3 0.01462766      3 0.6063830 0.7180851 0.05348973
4 0.01329787     11 0.4680851 0.7393617 0.05400635
5 0.01063830     14 0.4202128 0.7393617 0.05400635
6 0.01000000     15 0.4095745 0.7340426 0.05387913

Variable importance
         glucose_conc                   BMI  Diabetes_pedigree_fn                   Age X2.Hour_serum_insulin        blood_pressure  No.of_times_pregnant   skin_fold_thickness 
                   47                    13                     9                     8                     7                     7                     4                     4 

Node number 1: 538 observations,    complexity param=0.3351064
  predicted class=NO   expected loss=0.3494424  P(node) =1
    class counts:   350   188
   probabilities: 0.651 0.349 
  left son=2 (409 obs) right son=3 (129 obs)
  Primary splits:
      glucose_conc          < 143.5  to the left,  improve=52.88214, (0 missing)
      BMI                   < 29.55  to the left,  improve=24.45672, (0 missing)
      Age                   < 28.5   to the left,  improve=24.10181, (0 missing)
      No.of_times_pregnant  < 6.5    to the left,  improve=16.89834, (0 missing)
      X2.Hour_serum_insulin < 123.5  to the left,  improve=10.47522, (0 missing)
  Surrogate splits:
      X2.Hour_serum_insulin < 222.5  to the left,  agree=0.796, adj=0.147, (0 split)
```


glucose_conc< 143.5
glucose_conc< 101.5
NO
148/9
BMI< 26.3
NO
49/4
Age< 28.5
Diabetes_pedigree_fn< 0.503
BMI< 41.45
NO
48/5
YES
2/5
blood_pressure>=69
NO
14/3
YES
5/11
Age>=54.5
NO
9/2
blood_pressure>=89
NO
9/3
Diabetes_pedigree_fn< 0.7975
Age< 34.5
No.of_times_pregnant< 5.5
NO
14/5
YES
3/7
glucose_conc< 107
NO
5/3
YES
9/24
YES
2/11
glucose_conc< 161.5
Diabetes_pedigree_fn< 0.332
NO
15/4
YES
9/30
YES
9/62

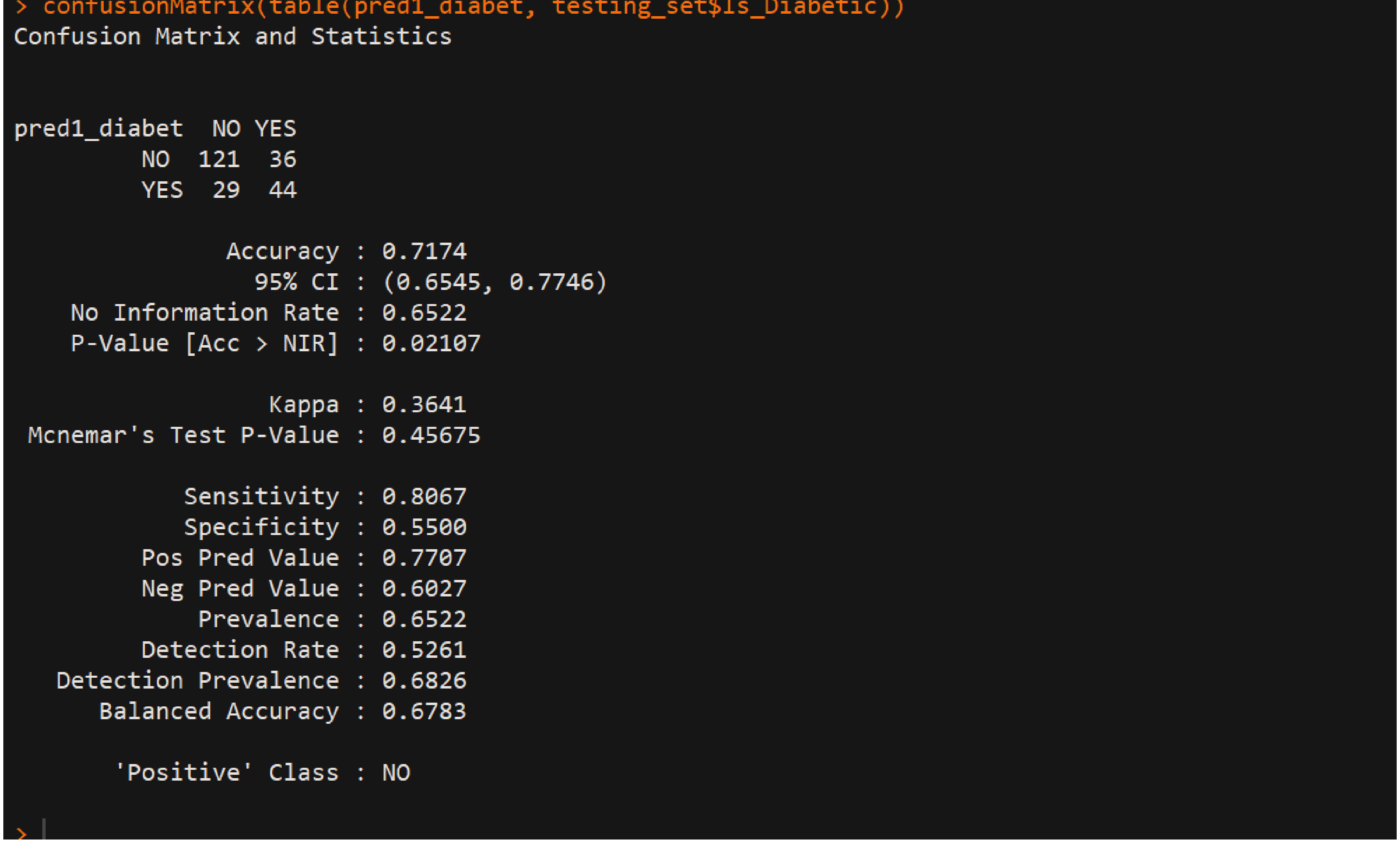


```
Confusion Matrix and Statistics


pred_diabet  NO YES
        NO  117  41
        YES  33  39

               Accuracy : 0.6783
                 95% CI : (0.6137, 0.7382)
    No Information Rate : 0.6522
    P-Value [Acc > NIR] : 0.2241

                  Kappa : 0.2739
 Mcnemar's Test P-Value : 0.4158

            Sensitivity : 0.7800
            Specificity : 0.4875
         Pos Pred Value : 0.7405
         Neg Pred Value : 0.5417
             Prevalence : 0.6522
         Detection Rate : 0.5087
   Detection Prevalence : 0.6870
      Balanced Accuracy : 0.6338

       'Positive' Class : NO

> 
```

Random Forest:

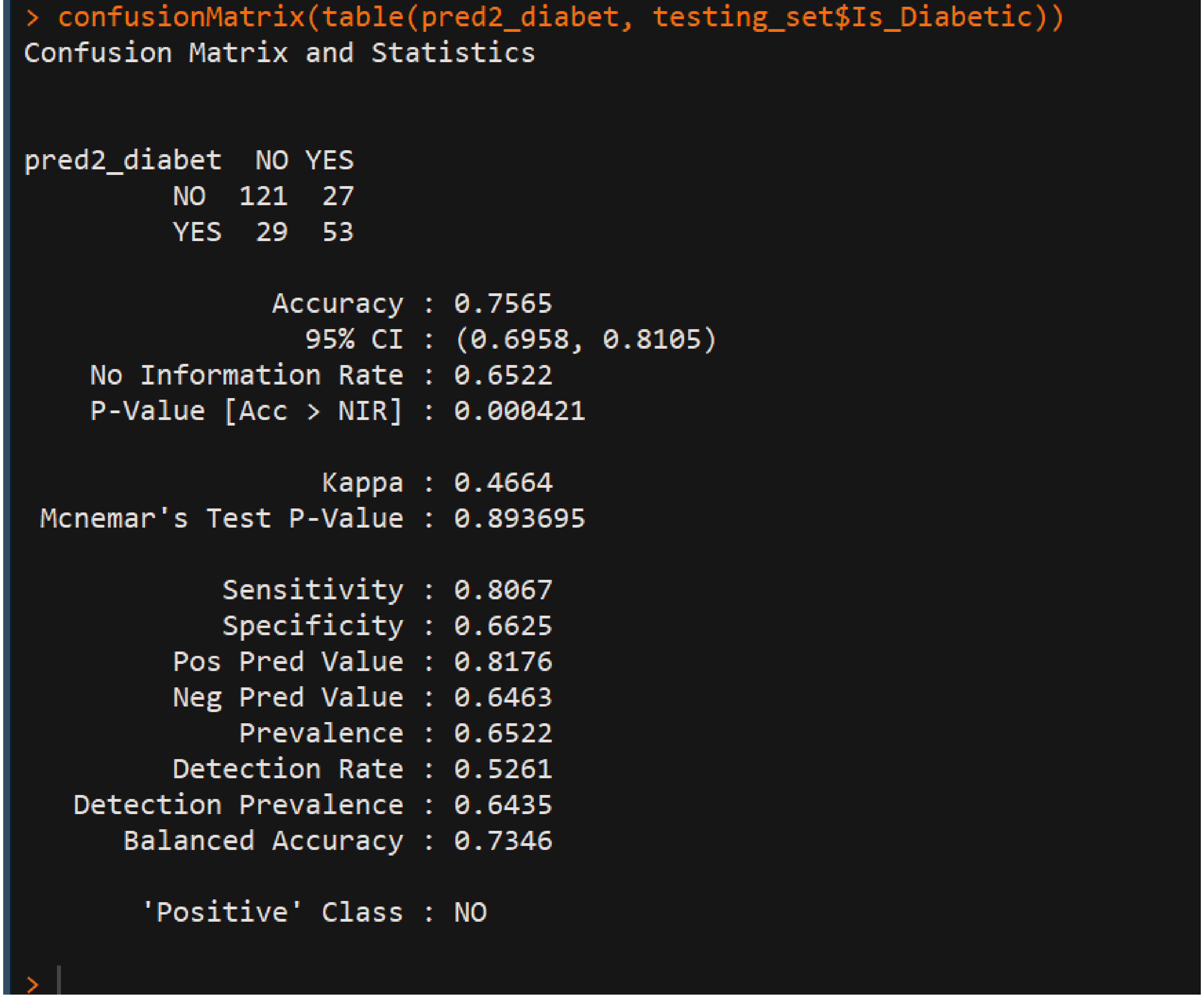


```
> diabet_forest <- randomForest(Is_Diabetic~., data = training_set)
> diabet_forest

Call:
 randomForest(formula = Is_Diabetic ~ ., data = training_set) 
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 23.61%
Confusion matrix:
     NO YES class.error
NO  299  51   0.1457143
YES  76 112   0.4042553
> 
```

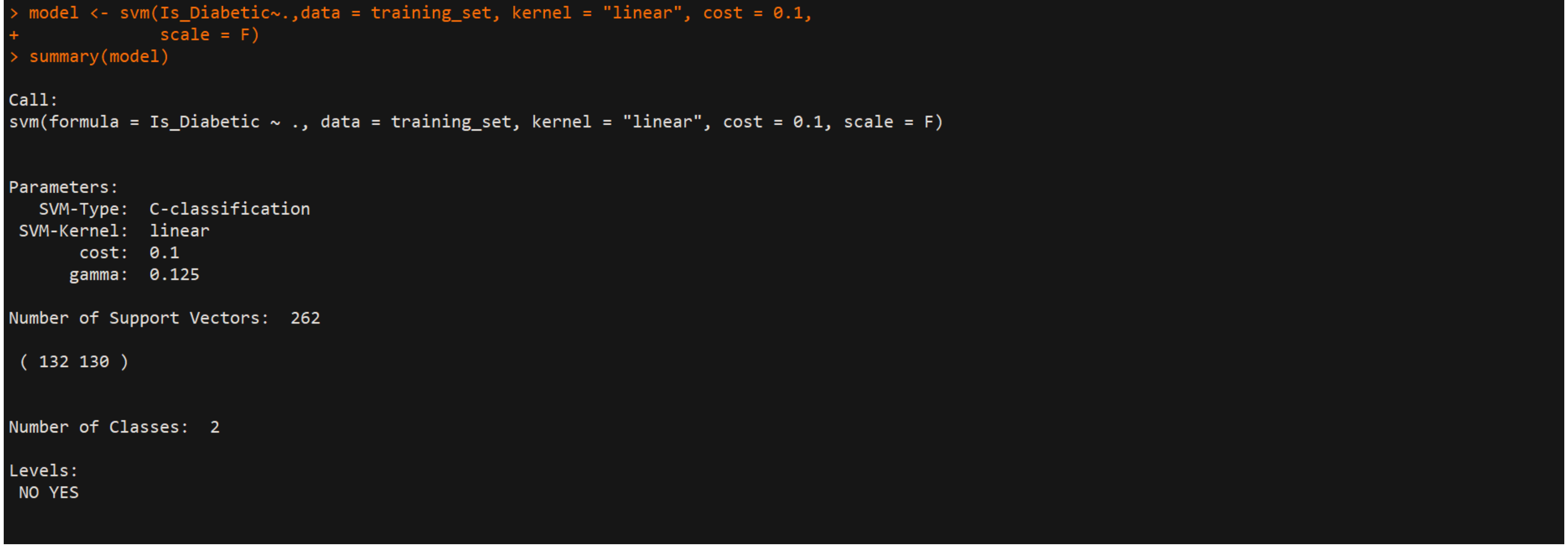


```
> confusionMatrix(table(pred1_diabet, testing_set$Is_Diabetic))
Confusion Matrix and Statistics


pred1_diabet  NO YES
         NO  121  36
         YES  29  44

               Accuracy : 0.7174
                 95% CI : (0.6545, 0.7746)
    No Information Rate : 0.6522
    P-Value [Acc > NIR] : 0.02107

                  Kappa : 0.3641
 Mcnemar's Test P-Value : 0.45675

            Sensitivity : 0.8067
            Specificity : 0.5500
         Pos Pred Value : 0.7707
         Neg Pred Value : 0.6027
             Prevalence : 0.6522
         Detection Rate : 0.5261
   Detection Prevalence : 0.6826
      Balanced Accuracy : 0.6783

       'Positive' Class : NO

> 
```

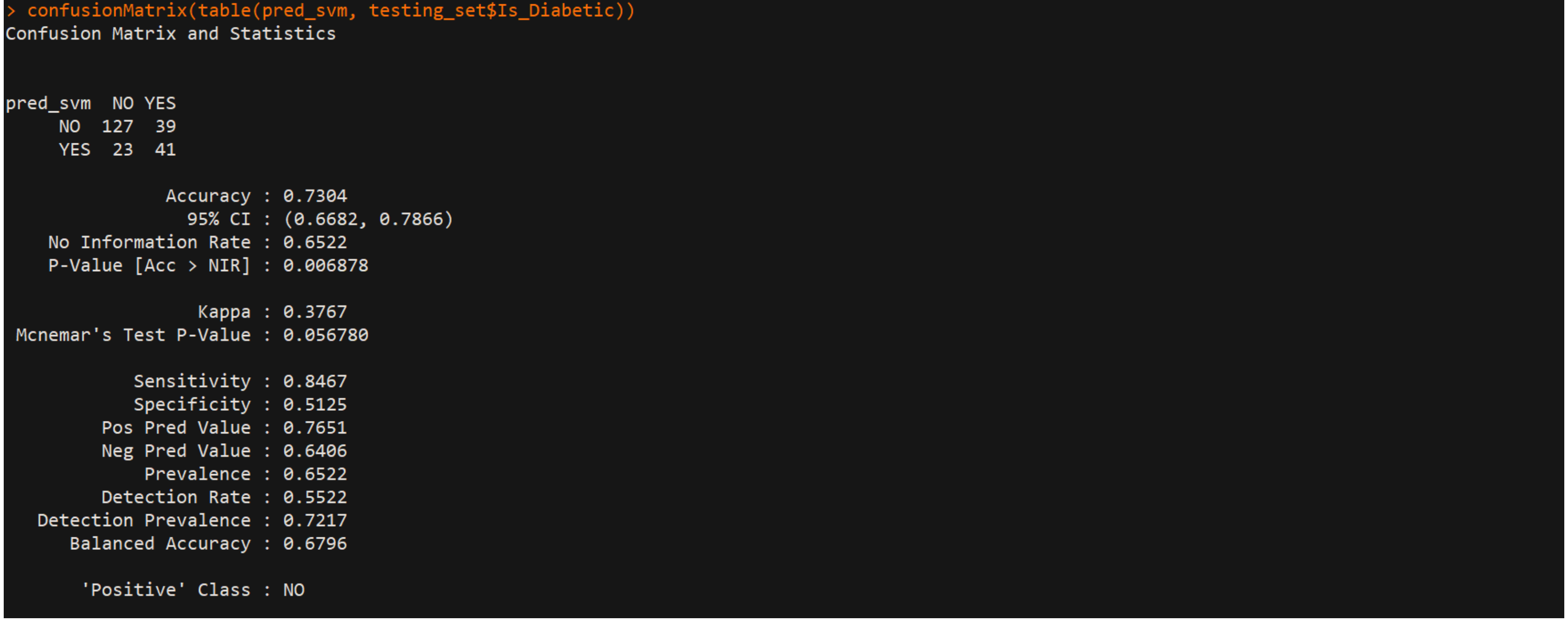
Naïve Bayes:



```
> confusionMatrix(table(pred2_diabet, testing_set$Is_Diabetic))
Confusion Matrix and Statistics


pred2_diabet  NO YES
         NO  121  27
         YES  29  53

               Accuracy : 0.7565
                 95% CI : (0.6958, 0.8105)
    No Information Rate : 0.6522
    P-Value [Acc > NIR] : 0.000421

                  Kappa : 0.4664
 Mcnemar's Test P-Value : 0.893695

            Sensitivity : 0.8067
            Specificity : 0.6625
         Pos Pred Value : 0.8176
         Neg Pred Value : 0.6463
             Prevalence : 0.6522
         Detection Rate : 0.5261
   Detection Prevalence : 0.6435
      Balanced Accuracy : 0.7346

       'Positive' Class : NO

> 
```



```
> model <- svm(Is_Diabetic~.,data = training_set, kernel = "linear", cost = 0.1,
+              scale = F)
> summary(model)

Call:
svm(formula = Is_Diabetic ~ ., data = training_set, kernel = "linear", cost = 0.1, scale = F)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  linear
       cost:  0.1
      gamma:  0.125

Number of Support Vectors:  262

 ( 132 130 )


Number of Classes:  2

Levels:
 NO YES
```



```
> confusionMatrix(table(pred_svm, testing_set$Is_Diabetic))
Confusion Matrix and Statistics


pred_svm  NO YES
     NO  127  39
     YES  23  41

               Accuracy : 0.7304
                 95% CI : (0.6682, 0.7866)
    No Information Rate : 0.6522
    P-Value [Acc > NIR] : 0.006878

                  Kappa : 0.3767
 Mcnemar's Test P-Value : 0.056780

            Sensitivity : 0.8467
            Specificity : 0.5125
         Pos Pred Value : 0.7651
         Neg Pred Value : 0.6406
             Prevalence : 0.6522
         Detection Rate : 0.5522
   Detection Prevalence : 0.7217
      Balanced Accuracy : 0.6796

       'Positive' Class : NO
```