R Notebook

Hide

```
library(tidyverse)
library(caret)
library(rpart)
library(knitr) #Dynamic Report Generator including use of LateX, HTML
library(gridExtra)
library(corrplot)
library(Boruta) #Feature selection
library(randomForest) #Random forest
library(ggRandomForests) #variable importance random forest
library(DMwR) #BINARY CLASSIFICATION
library(pROC) #ROC PLOT
library(shinydashboard)
library(shiny)
library(readxl)
library(plotly)
library(ROCR)
library(xgboost)
```

Hide

```
load("model.RData")
load("file.RData")
```

Hide

```
set.seed(1337)
Predictions_rf <- predict(model_rf, smote_test)
confusionMatrix(Predictions_rf, smote_test$Attrition)
```

```
Confusion Matrix and Statistics


          Reference
Prediction  No Yes
       No  240  57
```

```
       Yes  44 156

               Accuracy : 0.7968
                 95% CI : (0.7587, 0.8313)
    No Information Rate : 0.5714
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.5819
 Mcnemar's Test P-Value : 0.2325

            Sensitivity : 0.8451
            Specificity : 0.7324
         Pos Pred Value : 0.8081
         Neg Pred Value : 0.7800
             Prevalence : 0.5714
         Detection Rate : 0.4829
   Detection Prevalence : 0.5976
      Balanced Accuracy : 0.7887

       'Positive' Class : No
```

Hide

```
set.seed(1337)
Predictions_glm <- predict(model_glm, smote_test)
confusionMatrix(Predictions_glm, smote_test$Attrition)
```

```
Confusion Matrix and Statistics

          Reference
Prediction  No Yes
       No  230  71
       Yes  54 142

               Accuracy : 0.7485
                 95% CI : (0.7079, 0.7861)
```

```
    No Information Rate : 0.5714
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.4813
 Mcnemar's Test P-Value : 0.1524

            Sensitivity : 0.8099
            Specificity : 0.6667
         Pos Pred Value : 0.7641
         Neg Pred Value : 0.7245
             Prevalence : 0.5714
         Detection Rate : 0.4628
   Detection Prevalence : 0.6056
      Balanced Accuracy : 0.7383

       'Positive' Class : No
```

Hide

```
set.seed(1337)
Predictions_svm <- predict(model_svm,smote_test)
confusionMatrix(Predictions_svm, smote_test$Attrition)
```

```
Confusion Matrix and Statistics

          Reference
Prediction  No Yes
       No  242  77
       Yes  42 136

               Accuracy : 0.7606
                 95% CI : (0.7206, 0.7974)
    No Information Rate : 0.5714
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5009
```

```
 Mcnemar's Test P-Value : 0.001828

            Sensitivity : 0.8521
            Specificity : 0.6385
         Pos Pred Value : 0.7586
         Neg Pred Value : 0.7640
             Prevalence : 0.5714
         Detection Rate : 0.4869
   Detection Prevalence : 0.6419
      Balanced Accuracy : 0.7453

       'Positive' Class : No
```

Hide

```
Predictions_xgb <- predict(model_xgb, smote_test)
confusionMatrix(Predictions_xgb, smote_test$Attrition)
```

```
Confusion Matrix and Statistics

          Reference
Prediction  No Yes
       No  249  48
       Yes  35 165

               Accuracy : 0.833
                 95% CI : (0.7972, 0.8647)
    No Information Rate : 0.5714
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.6564
 Mcnemar's Test P-Value : 0.1878

            Sensitivity : 0.8768
            Specificity : 0.7746
         Pos Pred Value : 0.8384
```

```
         Neg Pred Value : 0.8250
             Prevalence : 0.5714
         Detection Rate : 0.5010
   Detection Prevalence : 0.5976
      Balanced Accuracy : 0.8257

       'Positive' Class : No
```

Hide

```
set.seed(1337)
Predictions_lda <- predict(model_lda,smote_test)
confusionMatrix(Predictions_lda, smote_test$Attrition)
```

```
Confusion Matrix and Statistics

          Reference
Prediction  No Yes
       No  229  72
       Yes  55 141

               Accuracy : 0.7445
                 95% CI : (0.7037, 0.7823)
    No Information Rate : 0.5714
    P-Value [Acc > NIR] : 7.739e-16

                  Kappa : 0.473
 Mcnemar's Test P-Value : 0.1557

            Sensitivity : 0.8063
            Specificity : 0.6620
         Pos Pred Value : 0.7608
         Neg Pred Value : 0.7194
             Prevalence : 0.5714
         Detection Rate : 0.4608
   Detection Prevalence : 0.6056
```

```
      Balanced Accuracy : 0.7342

       'Positive' Class : No
```

Hide

```
roc_rf <- roc(as.numeric(smote_test$Attrition), as.numeric(Predictions_rf))
roc_rf$auc
```

```
Area under the curve: 0.7887
```

Hide

```
roc_svm <- roc(as.numeric(smote_test$Attrition), as.numeric(Predictions_svm))
roc_svm$auc
```

```
Area under the curve: 0.7453
```

Hide

```
roc_xgb <- roc(as.numeric(smote_test$Attrition), as.numeric(Predictions_xgb))
roc_xgb$auc
```

```
Area under the curve: 0.8257
```

Hide

```
roc_lda <- roc(as.numeric(smote_test$Attrition), as.numeric(Predictions_lda))
roc_lda$auc
```

```
Area under the curve: 0.7342
```

Hide

```
roc_glm <- roc(as.numeric(smote_test$Attrition), as.numeric(Predictions_glm))
roc_glm$auc
```

```
Area under the curve: 0.7383
```

Hide

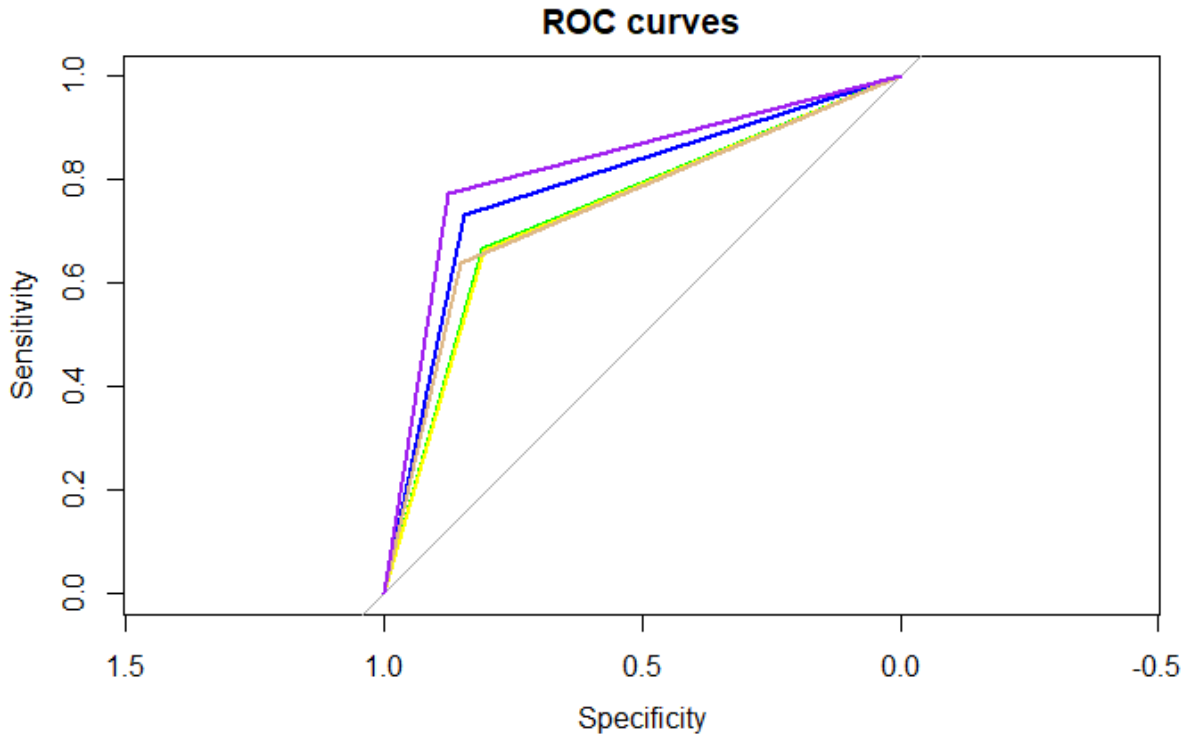
```
plot(roc_rf, ylim = c(0,1), main = "ROC curves", col = "blue")
plot(roc_glm, ylim = c(0,1),  col = "green", add = T)
```

Hide

```
plot(roc_lda, ylim = c(0,1),   col = "yellow", add = T)
plot(roc_svm, ylim = c(0,1),   col = "burlywood", add = T)
```

Hide

```
plot(roc_xgb, ylim = c(0,1),  col = "purple", add = T)
```



Hide

```
ggplot(smote_train,aes(Yearswithoutchange,fill=Attrition)) +
  geom_density(alpha=0.5)
```

```
```