

PROJECT: GERMAN CREDIT DATA

Credit Risk Assessment

Optimizing the Predictive model

Otimizando o Modelo preditivo

1. Working Directory

Configurando o diretório de trabalho

```
setwd("C:/Users/Utilizador/repos/Formacao_cientista_de_dados/big_data_analytics_R_microsoft_azure_machine_learning")  
getwd()
```

Observação: Este código contém comandos para filtrar e transformar os dados de aluguel de bikes,

Este código foi criado para executar tanto no Azure, quanto no RStudio;

Para executar no Azure, altere o valor da variavel Azure para TRUE;

Ou seja, se o valor for FALSE, o código sera executado no RStudio;

2. Weighted RandomForest Model

Modelo randomForest ponderado

O pacote C50 permite que você dê peso aos erros, construindo assim um resultado ponderado

```
install.packages("C50") library(C50)
```

Criando uma Cost Function

```
Cost_func <- matrix(c(0, 1.5, 1, 0), nrow = 2, dimnames = list(c("1", "2"), c("1", "2")))
```

Criando o Modelo

```
?randomForest ?C5.0
```

Cria o modelo

```
modelo_v2 <- C5.0(CreditStatus ~ CheckingAcctStat + Purpose + CreditHistory + SavingsBonds + Employment, data = dados_treino, trials = 100, cost = Cost_func)
print(modelo_v2)
```

Dataframes com valores observados e previstos

```
previsoes_v2 <- data.frame(observado = dados_teste$CreditStatus, previsto = predict(object = modelo_v2, newdata = dados_teste))
```

Calculando a Confusion Matrix em R (existem outras formas).

Label 1 - Credito Ruim

Label 2 - Credito Bom

Formulas

```
Accuracy <- function(x){ (x[1,1] + x[2,2]) / (x[1,1] + x[1,2] + x[2,1] + x[2,2]) }
Recall <- function(x){ x[1,1] / (x[1,1] + x[1,2]) }
Precision <- function(x){ x[1,1] / (x[1,1] + x[2,1]) }
W_Accuracy <- function(x){ (x[1,1] + x[2,2]) / (x[1,1] + 5 * x[1,2] + x[2,1] + x[2,2]) }
F1 <- function(x){ 2 * x[1,1] / (2 * x[1,1] + x[1,2] + x[2,1]) }
```

Criando a confusion matrix.

```
confMat_v2 <- matrix(unlist(Map(function(x, y){sum(ifelse(previsoes_v2[, 1] == x & previsoes_v2[, 2] == y, 1, 0) )}), c(2, 1, 2, 1), c(2, 2, 1, 1))), nrow = 2)
```

Criando um dataframe com as estatísticas dos testes

```
df_mat <- data.frame(Category = c("Credito Ruim", "Credito Bom"), Classificado_como_ruim =  
c(confMat_v2[1,1], confMat_v2[2,1]), Classificado_como_bom = c(confMat_v2[1,2], confMat_v2[2,2]), Ac-  
curacy_Recall = c(Accuracy(confMat_v2), Recall(confMat_v2)), Precision_WAcc = c(Precision(confMat_v2),  
W_Accuracy(confMat_v2)))  
print(df_mat)
```

Gerando Confusion Matrix com o Caret

```
library(caret) confusionMatrix(previsoes_v2observado,previsoes_v2previsto)
```