

PROJECT: GERMAN CREDIT DATA

Credit Risk Assessment

Predictive Model

Criando o Modelo Preditivo no R

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. Classification Model

Criar um modelo de classificação baseado em randomForest

```
library(randomForest)
```

Cross Tabulation

```
?table table(Credit$CreditStatus)
```

Funcao para gerar dados de treino e dados de teste

```
splitData <- function(dataframe, seed = NULL) { if (!is.null(seed)) set.seed(seed) index <- 1:nrow(dataframe)
trainindex <- sample(index, trunc(length(index)/2)) trainset <- dataframe[trainindex, ] testset <-
dataframe[-trainindex, ] list(trainset = trainset, testset = testset) }
```

Gerando dados de treino e de teste

```
splits <- splitData(Credit, seed = 808)
```

Separando os dados

```
dados_treino <- splits$trainset dados_teste <- splits$testset
```

Verificando o numero de linhas

```
nrow(dados_treino) nrow(dados_teste)
```

Construindo o modelo

```
modelo <- randomForest( CreditStatus ~ CheckingAcctStat + Duration_f + Purpose + CreditHistory +
SavingsBonds + Employment + CreditAmount_f, data = dados_treino, ntree = 100, nodesize = 10)
```

Imprimindo o resultado

```
print(modelo)
```