

PROJECT: GERMAN CREDIT DATA

Credit Risk Assessment

Feature Selection #Seleção de Variáveis #####
#####

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. Data Collection and Transformation

Variável que controla a execução do script

```
Azure <- FALSE
```

```
2.1 Data Collect if(Azure){ source("src/ClassTools.R") Credit <- maml.mapInputPort(1) }
```

3. Random Forest

Modelo randomForest para criar um plot de importância das variáveis

```
library(randomForest) modelo <- randomForest( CreditStatus ~ . - Duration - Age - CreditAmount -  
ForeignWorker - NumberDependents - Telephone - ExistingCreditsAtBank - PresentResidenceTime - Job -  
Housing - SexAndStatus - InstallmentRatePecnt - OtherDetorsGuarantors - Age_f - OtherInstalments, data  
= Credit, ntree = 100, nodesize = 10, importance = T)  
varImpPlot(modelo)  
outFrame <- serList(list(credit.model = modelo))
```

Output

```
if(Azure) maml.mapOutputPort("outFrame")
```