

# Analise Base de Dados de Crédito

*Luciano Teixeira*

*04 de novembro de 2018*

## Processamento Base de Dados de Crédito

O propósito deste relatório é analisar a base de dados de crédito com 2000 registros, identificando qual dos clientes, ao solicitar um empréstimo, possui a maior chance de efetuar o pagamento ou não deste empréstimo, levando em consideração:

**1 Idade**

**2 Renda**

**3 Histórico Financeiro**

Foi utilizado o seguinte algoritmo de Machine Learning para avaliar a mesma base:

### Árvore de Decisão

Para este algoritmo, foi adotado alguns procedimentos de ajustes, “ETL”, para eventuais correções, alterações, seja por conta de categorias de variáveis, seja por erro oriundos de intervenções manuais

**1 Importação da base de dados, com o propósito de ler o arquivo csv para sua classificação;**

**2 Eliminação da coluna de clientid, pois não há propósito de categoria ou classificação desta coluna;**

**3 Substituição de valores negativos pela média de idade positiva da base, a fim de minimizar a interferência nos dados;**

**4 Substituição de valores nulos “NA” pela média da idade positiva, a fim de minimizar a interferência nos dados;**

**5 Efetuado o nivelamento da escala, por exemplo, entre a idade e a renda, pois o valor da renda em escala comparado à idade, é muito maior, sendo assim, a aprendizagem não é eficiente;**

##6- O Encode da Classe ou transformação dos atributos categóricos em discretos, é fundamental pois diversas bibliotecas não aceitam como entrada, atributos categóricos. ##7- Divisão da base em dados de treinamento e dados de teste.

### Árvore de Decisão

#### Importando a Base de Dados

```
base = read.csv('BaseCredito.csv')
```

## Eliminando coluna clientid

```
base$clientid = NULL
```

## Preencher os valores negativos com a média dos valores positivos da coluna Age

```
base$age = ifelse(base$age < 0, 40.92, base$age)
```

## Preencher os valores nulos

```
base$age = ifelse(is.na(base$age), mean(base$age, na.rm = TRUE), base$age)
```