

DOUGLAS VINÍCIUS GONÇALVES ARAÚJO

**Modelo de Regressão Logística Aplicada a  
Previsão de Inadimplência sobre Cartão de  
Crédito de uma Instituição Financeira**

JI-PARANÁ

2022

DOUGLAS VINÍCIUS GONÇALVES ARAÚJO

**Modelo de Regressão Logística Aplicada a Previsão de  
Inadimplência sobre Cartão de Crédito de uma Instituição  
Financeira**

Relatório de Estágio Supervisionado apresentado como Trabalho de Pesquisa à Coordenação do Curso de Bacharelado em Estatística da Universidade Federal de Rondônia.

UNIVERSIDADE FEDERAL DE RONDÔNIA – UNIR  
DEPARTAMENTO DE MATEMÁTICA E ESTATÍSTICA  
RELATÓRIO DE PESQUISA

Jl-PARANÁ

2022

# Agradecimentos

Os agradecimentos...

*"Os livros servem para nos lembrar quanto somos estúpidos e tolos. São o guarda pretoriano de César, cochichando enquanto o desfile ruge pela avenida: Lembre-se, César, tu és mortal. A maioria de nós não pode sair correndo por aí, falar com todo mundo, conhecer todas as cidades do mundo, não temos tempo, dinheiro ou tantos amigos assim. As coisas que você está procurando, Montag, estão no mundo, mas a única possibilidade que o sujeito comum terá de ver noventa e nove por cento delas está num livro".*

*- Fahrenheit 451 de Ray Douglas Bradbury*

# Resumo

O objetivo deste trabalho é aplicar o modelo de regressão logística a dados de cartão de crédito de uma instituição financeira do estado de Rondônia.

**Palavras-chaves:** Risco de Crédito, Probabilidade de Default, Modelo de Regressão Logístico.

# Lista de ilustrações

Figura 1 – Machine Learning e suas aplicações . . . . .	11
---	----

## Lista de tabelas

# Lista de abreviaturas e siglas

ABNT	Associação Brasileira de Normas Técnicas
abnTeX	ABsurdas Normas para TeX



# Lista de símbolos

$\Gamma$	Letra grega Gama
$\Lambda$	Lambda
$\zeta$	Letra grega minúscula zeta
$f(x; \theta)$	Função de Densidade de Probabilidade
$\Pi$	Produtório

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
<b>1.1</b>	<b>Objetivos</b>	<b>10</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>11</b>
<b>2.1</b>	<b>Credit Scoring</b>	<b>11</b>
<b>2.2</b>	<b>Breve Introdução sobre Machine Learning</b>	<b>11</b>
<b>2.3</b>	<b>Modelo de Regressão Logística</b>	<b>11</b>
2.3.1	Interpretação dos Parâmetros	12
2.3.2	Estimação dos Parâmetros	12
2.3.3	Testes de Significância	12
2.3.4	Seleção de Variáveis	12
2.3.5	Desempenho dos Modelos	12
<b>3</b>	<b>METODOLOGIA E DADOS</b>	<b>13</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>14</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>15</b>
	<b>REFERÊNCIAS</b>	<b>16</b>
	<b>APÊNDICES</b>	<b>17</b>
	<b>APÊNDICE A – DESCRIÇÃO DAS VARIÁVEIS</b>	<b>18</b>
	<b>APÊNDICE B – SCRIPT R</b>	<b>19</b>

# 1 Introdução

## 1.1 Objetivos

O objetivo deste trabalho é desenvolver um modelo de previsão de risco de inadimplência dos tomadores de cartões de créditos de uma instituição Financeira do Estado de Rondônia.

## 2 REFERENCIAL TEÓRICO

### 2.1 Credit Scoring

### 2.2 Breve Introdução sobre Machine Learning

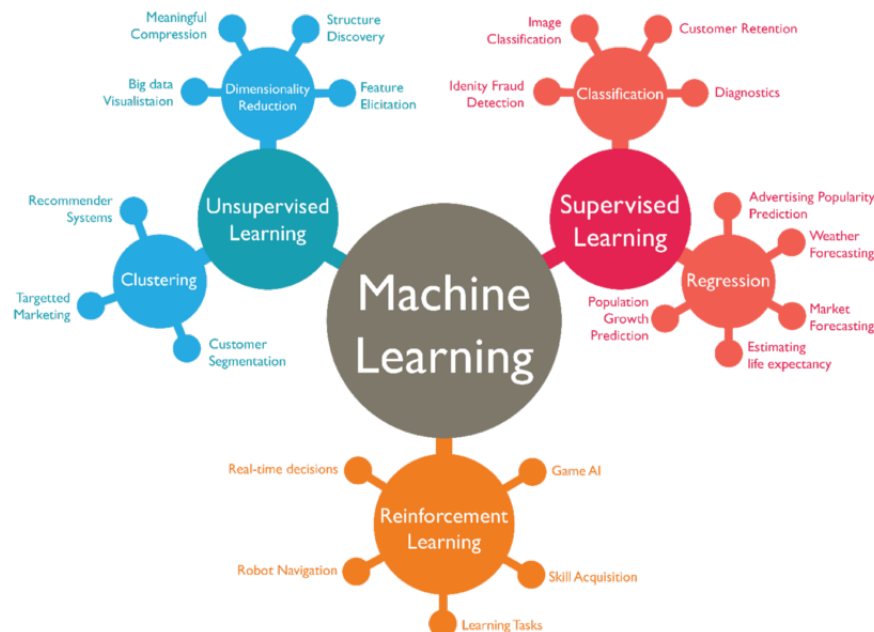
Uma definição básica sobre Machine Learning (Aprendizado de Máquina) é englobar um conjunto de regras com algoritmos e procedimentos que tem como objetivo de extrair informações a partir dos dados e dessas informações tomar uma decisão.

Segundo (GOODFELLOW; BENGIO; COURVILLE, 2016), os algoritmos de Machine Learning podem ser amplamente categorizados como aprendizados Supervisionado e Não-Supervisionados, sintetizando essa diferença, é o tipo de experiência durante o aprendizado do algoritmo.

### 2.3 Modelo de Regressão Logística

A regressão logística tem como principal uso modelar uma variável binária (0, 1), com base em mais variáveis, estas chamadas de variáveis explicativas ou preditoras. E

Figura 1 – Machine Learning e suas aplicações



Fonte: <https://becominghuman.ai/an-introduction-to-machine-learning>

comumente a variável resposta ou dependente, assim chama-se a variável binária do modelo. Conforme (HILBE, 2016), o melhor modelo ajustado aos dados é assumido que:

- Não há correlação entre as variáveis preditoras;
- Estejam significativamente relacionados com a resposta;
- Que as observações dos dados não estejam correlacionados com a resposta.

A resposta do modelo dito está conveniente a uma distribuição subjacente, ou seja, segue uma distribuição de Bernoulli. Como esta distribuição é um subconjunto da distribuição Binomial que a função de probabilidade pode ser expressa:

$$f(x; \theta) = \prod_{i=1}^n \theta_i^{x_i} (1 - \theta_i)^{1-x_i} \quad (2.1)$$

### 2.3.1 Interpretação dos Parâmetros

### 2.3.2 Estimação dos Parâmetros

### 2.3.3 Testes de Significância

### 2.3.4 Seleção de Variáveis

### 2.3.5 Desempenho dos Modelos

## 3 Metodologia e Dados

## 4 Resultados e Discussões

## 5 Considerações Finais



## Referências

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016. Citado na página [11](#).

HILBE, J. M. *Practical guide to logistic regression*. [S.l.]: crc Press, 2016. Citado na página [12](#).

## Apêndices



## APÊNDICE B – SCRIPT R