

DOUGLAS VINÍCIUS GONÇALVES ARAÚJO

Regressão Logística Aplicada a Análise de Inadimplência de Cartão de Crédito

JI-PARANÁ

2022

DOUGLAS VINÍCIUS GONÇALVES ARAÚJO

Regressão Logística Aplicada a Análise de Inadimplência de Cartão de Crédito

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

Lista de tabelas

Lista de abreviaturas e siglas

ABNT	Associação Brasileira de Normas Técnicas
abnTeX	ABsurdas Normas para TeX

Lista de símbolos

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta
Π	Produtório

Sumário

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

1 Introdução

1.1 Objetivos

2 REFERENCIAL TEÓRICO

2.1 Credit Scoring

2.2 Breve Introdução sobre Machine Learning

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 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

HILBE, J. M. *Practical guide to logistic regression*. [S.l.]: crc Press, 2016. Citado na página [11](#).

Apêndices

APÊNDICE A – DESCRIÇÃO DAS VARIÁVEIS

Coluna 1										Coluna 2	Coluna 3
Coluna 1	Coluna 2	Coluna 3	A	B	D	E	C	F	A	B	C
D										E	F

APÊNDICE B – SCRIPT R