

UNIVERSIDADE FEDERAL DE ALFENAS  
INSTITUTO DE CIÊNCIAS SOCIAIS APLICADAS

**HUMBERTO MARCILIO MARTINS**

**Previsão de risco de crédito no mercado de imóveis utilizando  
regressão logística ordinal, redes neurais artificiais e sistemas  
baseados em regras *fuzzy***

VARGINHA - MG

2020

**HUMBERTO MARCILIO MARTINS**

Previsão de risco de crédito no mercado de imóveis utilizando regressão logística ordinal, redes neurais artificiais e sistemas baseados em regras *fuzzy*

Trabalho apresentado como fase final do Curso de Pós-Graduação *Lato sensu* para a obtenção do grau de Especialista em Modelagem em Ciência e Tecnologia pela Universidade Federal de Alfenas.

Orientador: Leandro Ferreira

VARGINHA - MG

2020

**HUMBERTO MARCILIO MARTINS**

**Previsão de risco de crédito no mercado de imóveis utilizando regressão logística ordinal, redes neurais artificiais e sistemas baseados em regras *fuzzy***

A Banca examinadora abaixo-assinada, aprova a Monografia apresentada como parte dos requisitos para obtenção do grau de Especialista em Modelagem em Ciência e Tecnologia pelo Instituto de Ciências Sociais Aplicadas da UNIFAL-MG.

Trabalho aprovado em: 11 de novembro de 2020

Prof. Orientador Leandro Ferreira  
Instituição: UNIFAL-MG/Campus Varginha

Prof. Membro 1 Lázaro Eduardo da Silva  
Instituição: CEFET-MG/Campus Varginha

Prof. Membro 2 Silvio Antonio Bueno Salgado  
Instituição: UNIFAL-MG/Campus Varginha

VARGINHA - MG

2020

# RESUMO

O índice de inadimplentes vem aumentando em todo o mundo. No Brasil, a situação é um pouco mais preocupante, uma vez que o índice cresceu de forma mais ampla. Em 2018, o índice de inadimplência aumentou 4,41% em comparação com 2017. Entre as regiões do Brasil, o sudeste foi a que mais aumentou a inadimplência no primeiro semestre de 2019, sendo um aumento de 3,4% em comparação com o mesmo período do ano de 2018. Para suprir essas dificuldades, o mercado utiliza alguns modelos estatísticos e computacionais para classificar clientes inadimplentes ou não. Dentre os modelos utilizados, pode-se citar o uso da regressão logística, redes neurais artificiais e sistemas baseados em regras *fuzzy*. Este trabalho tem como principal objetivo, utilizar regressão logística ordinal, redes neurais artificiais e sistemas baseados em regras *fuzzy* para previsão de risco de crédito de uma carteira de clientes em uma loteadora do sul de Minas Gerais.

**Palavras-chave:** *Inadimplência, regressão logística ordinal, redes neurais artificiais, sistemas baseados em regras fuzzy*

# ABSTRACT

The rate of defaulters has been increasing worldwide. In Brazil, the situation is a little more worrying, since the index has grown more widely. In 2018, the default rate increased 4.41 % compared to 2017. Among the regions of Brazil, the southeast was the one with the highest increase in default in the first half of 2019, an increase of 3.4 % in comparison with the same period in 2018. To address these difficulties, the market uses some statistical and computational models to classify customers who are in default or not. Among the models used, we can mention the use of logistic regression, artificial neural networks and fuzzy rule-based systems. This work has as main objective, to use ordinal logistic regression, artificial neural networks and fuzzy rule-based systems for predicting the credit risk of a client portfolio in a land development company in the south of Minas Gerais.

**Keywords:** *financial default, ordinal logistic regression, artificial neural networks, fuzzy rule-based systems*

# Lista de tabelas

Tabela 1 – Arquiteturas das redes neurais. . . . .	17
Tabela 2 – Sistemas baseados em regras <i>fuzzy</i> propostos. . . . .	18
Tabela 3 – Informações dos pacotes utilizados no <i>software R</i> . . . . .	18
Tabela 4 – Variáveis agrupadas por faixas. . . . .	19
Tabela 5 – Resumo das variáveis. . . . .	19
Tabela 6 – Médias das porcentagens de acerto e média de erros nas etapas de treinamento e teste - redes neurais artificiais. . . . .	20
Tabela 7 – Médias das porcentagens de acerto nas etapas de treinamento e teste - sistema baseado em regras <i>fuzzy</i> . . . . .	20

# Lista de ilustrações

Figura 1 – Rede <i>feedforward</i> de camadas múltiplas. Fonte: Silva, Spatti e Flauzino (2010). . . . .	13
Figura 2 – Esquema ilustrativo de um SBRF. . . . .	14

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>8</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> . . . . .	<b>10</b>
<b>2.1</b>	<b>Risco de crédito</b> . . . . .	<b>10</b>
<b>2.2</b>	<b>Regressão logística ordinal</b> . . . . .	<b>11</b>
<b>2.3</b>	<b>Rede neural artificial</b> . . . . .	<b>12</b>
<b>2.4</b>	<b>Sistema baseado em regras <i>fuzzy</i></b> . . . . .	<b>14</b>
<b>3</b>	<b>METODOLOGIA</b> . . . . .	<b>16</b>
<b>3.1</b>	<b>Dados</b> . . . . .	<b>16</b>
<b>3.2</b>	<b>Regressão logística ordinal</b> . . . . .	<b>16</b>
<b>3.3</b>	<b>Rede neural artificial</b> . . . . .	<b>16</b>
<b>3.4</b>	<b>Sistema baseado em regras <i>fuzzy</i></b> . . . . .	<b>17</b>
<b>3.5</b>	<b>Implementação computacional</b> . . . . .	<b>18</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	<b>19</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b> . . . . .	<b>21</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>22</b>

# 1 INTRODUÇÃO

O índice de inadimplência vem aumentando em todo o mundo, sendo que no Brasil, o índice cresceu de forma mais acentuada. Em 2018, o índice de inadimplentes aumentou 4,41% em comparação a 2017 (BEZERRA, 2019). Em 2019, o número de pessoas inadimplentes no Brasil foi maior que o número de habitantes dos países da América Latina, exceto do México (ROCHA, 2019). Segundo dados da Confederação Nacional de Dirigentes Lojistas (CDNL, 2019), a região do sudeste brasileiro foi a que mais aumentou a inadimplência no primeiro semestre de 2019, sendo um aumento de 3,4% em comparação com o mesmo período do ano de 2018.

Saber lidar com a inadimplência tem sido um dos desafios das pequenas e médias empresas, uma vez que o seu capital de giro é menor, podendo ter maiores dificuldades financeiras (LUCENA et al., 2015). Com o advento da informática nos anos 90, o mercado de risco passou a utilizar modelos computacionais para lidar com a inadimplência (SELAU; RIBEIRO, 2009). Embora a utilização de modelos computacionais não tem custo elevado (VALENTE et al., 2014), tais modelos raramente são aplicados em pequenas e médias empresas de imóveis.

Dentre os modelos computacionais, tem-se as redes neurais artificiais e os sistemas baseados em regras *fuzzy*, utilizados na previsão de risco de crédito. As redes neurais artificiais têm sido amplamente utilizadas para classificação em diversas áreas, como biologia, engenharia e finanças (WUERGES; BORBA, 2010). Modelos baseados na teoria dos conjuntos *fuzzy* têm sido utilizados na análise de crédito (SYAU; HSIEH; LEE, 2001) e na avaliação de processos de insolvência, alocação de recursos e gestão de riscos (BORBA; DILL, 2007). Ainda para a previsão de risco de crédito, modelos estatísticos podem ser utilizados, como o modelo de regressão logística, para a classificação de clientes como adimplentes ou inadimplentes (GONÇALVES; GOUVÊA; MANTOVANI, 2013).

O presente trabalho tem como principal objetivo apresentar modelos baseados em regressão logística ordinal, redes neurais artificiais e sistemas baseados em regras *fuzzy*, para previsão de risco de crédito de uma carteira de clientes de uma loteadora do sul de Minas Gerais. Além disso, este trabalho tem como objetivos secundários, contribuir para as empresas de imóveis de médio e pequeno porte com um modelo de classificação de baixo custo e incrementar a linha de pesquisa com uso de modelos computacionais. Por fim, vale ressaltar que a comparação dos modelos não faz parte dos objetivos propostos.

A seguir, será apresentado o referencial teórico com as definições de risco de crédito, regressão logística ordinal, redes neurais artificiais e sistemas baseados em regras *fuzzy*. Será apresentada a metodologia contendo informações dos dados utilizados e como foi aplicados cada um dos modelos computacionais definidos no referencial teórico. Por fim, serão apresen-

tados os resultados e discussões, seguidos pelas considerações finais.

## 2 REFERENCIAL TEÓRICO

### 2.1 Risco de crédito

O crédito começou a se tornar um negócio cada vez mais rentável no Brasil a partir de 1994 com o plano real e a estabilidade da moeda brasileira (CAMARGOS; ARAÚJO; CAMARGOS, 2012). Segundo Brito e Assaf Neto (2008), o crédito pode ter diversas interpretações. No que tange as instituições financeiras, trata-se de um empréstimo ou financiamento para um tomador de bens ou valores que serão pagos em uma futura data. Schrickel (1995) traz que o crédito pode ser definido como toda ação de alguém emprestar ou ceder por um determinado tempo, parte de seus bens ou valores a um terceiro, com a esperança de obter os valores ou o patrimônio emprestado de forma integral após o tempo acordado.

Segundo Palmuti e Picchiali (2012), o crédito pode ser entendido de uma forma mais abrangente como uma ferramenta que tem a finalidade de financiar agentes econômicos de uma nação em suas necessidades de consumo e investimento, contribuindo para o desenvolvimento da economia.

O crédito imobiliário no Brasil cresceu a partir dos meados de 2005 devido a mudanças nas leis, fomento na construção civil e controle da inflação, com predominância dos bancos privados (MACHADO; CERETTA; VIEIRA, 2014). De acordo com Mendonça (2013), com a crise de 2008, os bancos públicos passaram a ser os maiores financiadores devido ao incentivo do governo, principalmente por conta do programa "Minha casa, Minha vida", que concedeu crédito a população de baixa renda. Uma das principais leis que permitiu o fomento do crédito imobiliário foi a alienação fiduciária do bem imóvel, consolidada em agosto de 2004, que permitiu um contrato de compra e venda, onde o comprador toma posse do imóvel somente após a quitação do contrato. Tal lei sustentou empresas incorporadoras e os financiamentos dos bancos públicos, tendo o imóvel sempre como a garantia do contrato, reduzindo portanto o risco do crédito.

Pelo fato do crédito envolver a esperança de obter o retorno dos patrimônios, todo crédito possui um certo risco (possibilidade do retorno não acontecer de forma integral) (SCHRICKEL, 1995). Dessa forma, em qualquer operação de concessão de crédito, o risco de crédito vai estar presente, fazendo parte do dia a dia das organizações de gestão (CAMARGOS et al., 2010). Para Brito e Assaf Neto (2008), o risco de crédito pode ser definido como probabilidade do tomador não liquidar suas obrigações da forma pactuada com o credor. Para Camargos et al. (2010), o risco de crédito é a possibilidade do tomador não cumprir com as expectativas do credor em ter o retorno conforme o combinado, tornando o tomador como um devedor classificado como inadimplente.

Visto que a todo crédito concedido tem-se um risco de crédito associado, mensurar o risco de crédito é de grande importância para os credores. A gestão do risco de crédito possibilita baixar o nível de incertezas e perdas dos credores (CAMARGOS et al., 2010). Nesse ponto, os modelos estatísticos e computacionais têm sido importantes ferramentas de análises na gestão de risco de crédito, podendo classificar clientes em bons e maus pagadores (GONÇALVES; GOUVÊA; MANTOVANI, 2013).

Gonçalves, Gouvêa e Mantovani (2013) realizaram uma pesquisa na área de previsão de risco de crédito de uma instituição financeira utilizando o modelo de regressão logística. Neste trabalho, os autores dividiram a validação do modelo em duas categorias, sendo uma para bons pagadores e outra para maus pagadores. Os resultados foram 72% de acerto para maus pagadores e 70,4% de acerto para bons pagadores. Os autores consideraram os resultados satisfatórios, pois previsões acima de 65% já podem ser consideradas como bons resultados.

Camargos, Araújo e Camargos (2012) realizaram um trabalho utilizando o método de regressão logística hierárquica, com o objetivo de propor um modelo econométrico para uma instituição financeira pública de Minas Gerais. A intenção foi classificar as empresas em adimplentes e inadimplentes, considerando vinte e duas variáveis independentes, sendo que o modelo classificou corretamente, 85,5% das empresas.

Steiner et al. (2007) utilizaram modelos de redes neurais para previsão de crédito na rede bancária obtendo mais de 89% de acertos, uma taxa que os autores consideraram satisfatória, sendo que utilizaram o algoritmo *neurorule* do *software* WEKA (*Waikato Environment for Knowledge Analysis*). Selau e Ribeiro (2009) trabalharam com três modelos para previsão de crédito de uma rede de farmácias com crédito próprio, sendo dois modelos de estatística (regressão logística e análise discriminante) e um modelo de redes neurais. Os três modelos utilizados atingiram mais de 70% de acerto, sendo que o modelo de redes neurais apresentou melhor resultado.

Syau, Hsieh e Lee (2001) utilizaram a teoria dos conjuntos *fuzzy* para avaliar as incertezas que influenciam na análise de crédito. Freitas, Tavares e Montini (2008) apresentaram um modelo para avaliar o risco de crédito pessoal utilizando sistemas baseados em regras *fuzzy*. Os sistemas baseados em regras *fuzzy* apresentaram resultados superiores ao sistema de *credit scoring* tradicional, que é um modelo estatístico, baseado em experiências passadas.

## 2.2 Regressão logística ordinal

O modelo de regressão logística ordinal é uma extensão do modelo de regressão logística, que é um caso particular de modelos lineares generalizados (NELDER; WEDDERBURN, 1972; AGRESTI, 2013).

Considere  $Y$  como uma variável resposta binária, sendo que  $Y \sim \text{Bernoulli}(\pi)$ , com

$P(Y = 1) = \pi$  que é a probabilidade de sucesso e  $P(Y = 0) = 1 - \pi$  é a probabilidade de fracasso. Considerando uma variável explanatória  $x$ , a probabilidade de sucesso é dada por  $P(Y = 1|x) = \pi(x)$ . O modelo de regressão logística é dado da seguinte maneira:

$$\text{logito}[\pi(x)] = \ln \frac{\pi(x)}{1 - \pi(x)} = \eta = \beta_0 + \beta_1 x, \quad (2.1)$$

em que  $\text{logito}[\pi(x)]$  é a função de ligação do modelo, que liga o componente aleatório  $E(Y) = \pi$  ao componente sistemático linear  $\eta = \beta_0 + \beta_1 x$ , sendo que  $\beta_0$  e  $\beta_1$  são parâmetros desconhecidos a serem estimados.

O modelo de regressão logística ordinal é adequado quando a variável resposta ultrapassa duas categorias e são ordenadas. Os principais modelos de regressão logística ordinais são: modelo logito cumulativo, modelo logito categorias adjacentes e modelo logito razão contínua (AGRESTI, 2010). A seguir, é apresentado o modelo logito cumulativo.

Considere uma variável aleatória  $Y$  com  $c$  categorias ordinais, o logito cumulativo, incorporando um vetor de variáveis explanatórias ( $\mathbf{x}$ ), é definido por:

$$\text{logito}[P(Y \leq j|\mathbf{x})] = \ln \frac{P(Y \leq j|\mathbf{x})}{1 - P(Y \leq j|\mathbf{x})} = \ln \frac{\pi_1(\mathbf{x}) + \dots + \pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x}) + \dots + \pi_c(\mathbf{x})}, j = 1, \dots, c - 1, \quad (2.2)$$

em que  $j$  corresponde a categoria ordenada da variável resposta e  $\pi_j(\mathbf{x}) = P(Y = j|\mathbf{x})$  representa a probabilidade de ocorrência da  $j$ -ésima categoria da variável resposta para um dado vetor de variáveis explanatórias  $\mathbf{x}$ .

## 2.3 Rede neural artificial

As redes neurais artificiais (RNAs) são modelos computacionais baseados no funcionamento do cérebro humano, sendo capazes de reconhecer padrões através de experiências. Para a modelagem via redes neurais, é necessário definir o conjunto de dados a ser utilizado para o aprendizado da rede e para avaliar a capacidade de generalização. Além disso, é necessário escolher uma arquitetura apropriada e um algoritmo de treinamento.

A arquitetura de uma RNA pode possuir uma camada de entrada, uma ou mais camadas neurais intermediárias e uma camada neural de saída. A camada de entrada é responsável por receber todas as informações advindas do meio externo, sendo que nela não há neurônios. As intermediárias ou escondidas são aquelas que possuem a função de extrair as características referentes ao que está sendo analisado. Por fim, a camada neural de saída tem a capacidade de mostrar os resultados gerados pela rede (SILVA; SPATTI; FLAUZINO, 2010).

Existem vários tipos de arquiteturas de redes neurais artificiais, como por exemplo, a rede *feedforward*. Pela rede *feedforward*, conforme ilustrado na Figura 1, o fluxo de infor-

mações sempre será o mesmo, da esquerda para a direita, ou seja, da camada de entrada para a camada neural de saída. A Figura 1 apresenta uma rede *feedforward* de camadas múltiplas com  $n$  entradas na camada de entrada, duas camadas neurais escondidas com  $n_1$  neurônios na primeira camada e  $n_2$  neurônios na segunda camada e,  $m$  neurônios na camada neural de saída representando os respectivos valores de saída.

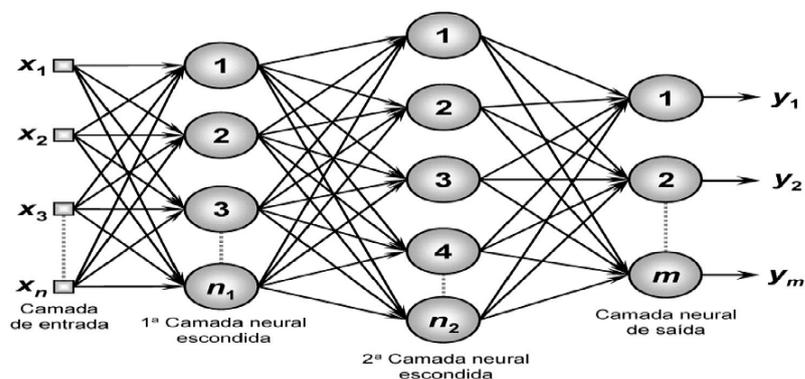


Figura 1 – Rede *feedforward* de camadas múltiplas.

Fonte: Silva, Spatti e Flauzino (2010).

Resumidamente, o processamento dos dados entre os neurônios é realizado considerando um combinador linear e uma função de ativação. O combinador linear desempenha o papel de unir todas as entradas em um único valor, por meio de uma soma ponderada. Cada entrada é multiplicada por seu respectivo peso e esses produtos são somados, chegando-se a um único valor, representando a soma total das entradas ponderadas por seus respectivos pesos. A função de ativação tem como objetivo limitar a saída do neurônio dentro de um intervalo. Existem vários tipos de função de ativação, como por exemplo, função sigmóide ou logística, linear e exponencial (CORRAR; PAULO; FILHO, 2009).

No processo de aprendizagem, uma RNA passará por uma etapa de treinamento para que consiga armazenar características que posteriormente serão utilizadas, por exemplo, para previsão. Pelo treinamento supervisionado, existe a necessidade de um conjunto de dados de entrada e saída. Os neurônios irão armazenar as características dos dados de entrada, sendo que os pesos serão reajustados via processo iterativo com o objetivo de minimizar o erro, ou seja, a diferença entre a saída da rede e a saída desejada. Como por exemplo, no treinamento supervisionado de uma rede *Perceptron* de múltiplas camadas, utiliza-se o algoritmo de retropropagação do erro (*backpropagation*). Pelo treinamento não-supervisionado, não existem dados de saída. No caso, a rede se auto-organiza pelas semelhanças existentes entre os dados de entrada. O conjunto de dados disponível é dividido em dois subconjuntos denominados de subconjuntos de treinamento e de teste. Cerca de 60 a 90% do total dos dados serão utilizados na etapa de treinamento. Para realização do teste, serão utilizados 40 a 10% dos dados. O processo de aprendizagem será interrompido quando o erro for minimizado e quando a rede for capaz de apresentar uma boa capacidade de generalização (SILVA; SPATTI; FLAUZINO, 2010).

## 2.4 Sistema baseado em regras *fuzzy*

A teoria de conjuntos *fuzzy* foi inicialmente introduzida por Zadeh (1965), tendo como objetivo, trabalhar informações imprecisas.

De acordo com Tanaka (1997), um conjunto *fuzzy*  $A$  de um conjunto universo  $U$  é definido por uma função de pertinência  $\mu_A$  representada por  $\mu_A : U \rightarrow [0, 1]$ , na qual  $\mu_A(x) \in [0, 1]$  é o grau de pertinência com que o elemento  $x$  de  $U$  pertence ao conjunto *fuzzy*  $A$ .

De acordo com Barros e Bassanezi (2010), um sistema baseado em regras *fuzzy* (SBRF) é constituído pela etapa de fuzzificação, que traduz as variáveis de entrada em conjuntos *fuzzy*, pela etapa de inferência, que realiza o raciocínio *fuzzy* com base num sistema de regras, que relaciona as variáveis de entrada com as de saída, e pela etapa de defuzzificação, que é a tradução da saída num valor numérico. A Figura 2 apresenta o esquema ilustrativo de um sistema baseado em regras *fuzzy*.

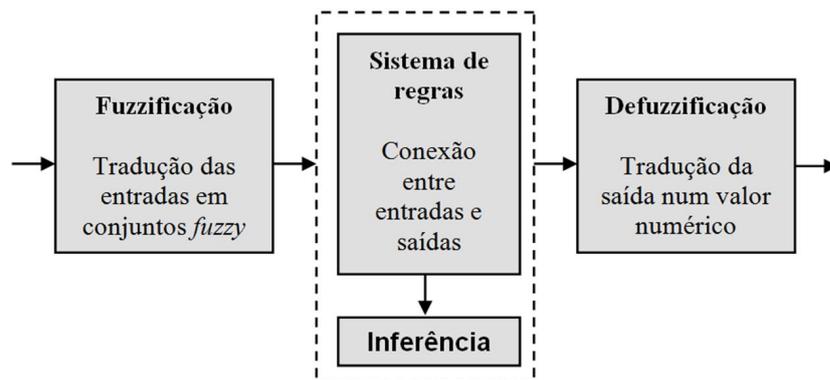


Figura 2 – Esquema ilustrativo de um SBRF.

O SBRF pode ser construído com base em banco de dados e opiniões de especialistas. Cada regra é composta por conectivos lógicos (se, e, ou, então) e pelas partes antecedente e consequente, por exemplo, *SE x é A E y é B ENTÃO z é C*, em que,  $A$ ,  $B$  e  $C$  são conjuntos *fuzzy*,  $x$  e  $y$  são as variáveis de entrada,  $z$  é a variável de saída, *SE x é A E y é B* é a parte antecedente e, *ENTÃO z é C*, é a parte consequente.

Um dos métodos de inferência mais utilizado é o tipo Mamdani, que aborda em cada regra o operador lógico E (operador mínimo) e agrega as regras por meio do operador lógico OU (operador máximo). Além do método de inferência tipo Mamdani, tem-se o método de Takagi-Sugeno, que está associado a um conjunto de regras que apresenta funções lineares nos consequentes de cada regra.

A defuzzificação consiste em traduzir um conjunto *fuzzy* em um número real. Existem diversos métodos de defuzzificação, sendo que o mais utilizado é o método do Centro de Gravidade, também chamado de Centróide ou Centro de Área.

O método do Centro de Gravidade (G) é semelhante à média ponderada para distribui-

ção de dados, com a diferença de que os pesos são os valores  $\mu_A(x)$ , que indicam o grau de pertinência do valor  $x_i$  no conjunto *fuzzy*  $A$  (BARROS; BASSANEZI, 2010).

Para um domínio discreto, tem-se:

$$G(A) = \frac{\sum_{i=0}^n x_i \mu_A(x_i)}{\sum_{i=0}^n \mu_A(x_i)} \quad (2.3)$$

Para um domínio contínuo, tem-se:

$$G(A) = \frac{\int_R x \mu_A(x) dx}{\int_R \mu_A(x) dx} \quad (2.4)$$

em que  $R$  é a região de integração.

## 3 METODOLOGIA

### 3.1 Dados

O banco de dados utilizado no presente trabalho refere-se a uma empresa de loteamentos de médio porte do sul de Minas Gerais. Os dados são de 216 clientes, sendo que os clientes foram classificados em risco baixo (clientes com uma parcela em atraso), risco moderado (clientes com duas parcelas em atraso) e risco alto (clientes com mais de duas parcelas em atraso).

O banco de dados é composto pelas variáveis: dias de atraso referentes ao pagamento das parcelas, que corresponde aos dias de atraso durante a presença do cliente na carteira; Índice Geral de Preços do Mercado - IGPM (%); idade do cliente (anos); número de parcelas contratadas; valor de entrada (R\$) e; taxa de desemprego (%).

O banco de dados foi dividido em dois conjuntos de dados, referentes a etapa de treinamento, que corresponde a 60% dos dados originais, e a etapa de teste, que corresponde a 40% dos dados originais.

### 3.2 Regressão logística ordinal

Foram construídos modelos logito cumulativos com chances proporcionais (MLC). Para tanto, foi considerada a variável aleatória  $Y$  representando o nível de risco do cliente, com 3 categorias ordinais (risco baixo, risco moderado e risco alto). Os modelos logito cumulativos com chances proporcionais incorporaram 6 variáveis explanatórias (dias de atraso, IGPM, idade do cliente, número de parcelas contratadas, valor de entrada e taxa de desemprego), sendo definidos por:

$$\text{logito}[P(Y \leq j|\mathbf{x})] = \beta_{j0} + \boldsymbol{\beta}' \mathbf{x} = \beta_{j0} + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_6 x_6, j = 1, 2, 3, \quad (3.1)$$

em que  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_6)'$  é um vetor de 6 parâmetros desconhecidos a serem estimados,  $\mathbf{x}$  representa o vetor dos valores das variáveis explanatórias. O modelo logito cumulativo tem para cada logito cumulativo ( $j = 1, 2, 3$ ) um intercepto  $\beta_{j0}$  e o mesmo coeficiente angular  $\boldsymbol{\beta}$ .

### 3.3 Rede neural artificial

Foram utilizadas seis arquiteturas de redes neurais artificiais, considerando diferentes quantidades de camadas escondidas e diferentes números de neurônios em cada camada,

conforme Tabela 1.

Tabela 1 – Arquiteturas das redes neurais.

Arquitetura	Número de camadas escondidas	Número de neurônios nas camadas escondidas
1	1	5
2	1	15
3	1	25
4	2	5 - 2
5	2	15 - 5
6	2	25 - 10

Fonte: Elaboração própria do autor.

As características das diferentes arquiteturas podem ser descritas da seguinte maneira:

- camada de entrada: foram consideradas 6 variáveis de entrada (dias de atraso, IGPM, idade do cliente, número de parcelas contratadas, valor de entrada e taxa de desemprego);
- camada de saída: 1 variável de saída (nível de risco do cliente);
- algoritmo de treinamento: RPROP–*Resilient Propagation* (RIEDMILLER; BRAUN, 1993). Tal algoritmo é uma extensão do algoritmo *backpropagation*, com uma adaptação direta no ajuste dos pesos da rede neural com base em informações locais durante o processamento;
- função de ativação (camadas escondida e saída): logística.

### 3.4 Sistema baseado em regras *fuzzy*

Foram gerados 9 tipos de sistemas baseados em regras *fuzzy* (SBRFs), considerando 6 variáveis de entrada (dias de atraso, IGPM, idade do cliente, número de parcelas contratadas, valor de entrada e taxa de desemprego) e uma variável de saída (nível de risco do cliente). Para tanto, foram considerados diferentes funções de pertinência e métodos de defuzzificação. As características dos diferentes SBRFs podem ser descritas da seguinte maneira:

- fuzzificação: admitiu-se 5, 25 e 45 conjuntos *fuzzy* para as funções de pertinência dos tipos triangular, trapezoidal e gaussiana;
- modelo de inferência do tipo Mamdani;
- defuzzificação: foi utilizado o método de defuzzificação centro de gravidade;
- algoritmo de aprendizado: foi utilizado o algoritmo de aprendizado WM (WANG; MENDEL, 1992), baseado na divisão dos domínios das variáveis de entrada e saída para obter os parâmetros das funções de pertinência.

A Tabela 2 apresenta os 9 SBRFs propostos, considerando diferentes funções de pertinência e quantidade de conjuntos *fuzzy*.

Tabela 2 – Sistemas baseados em regras *fuzzy* propostos.

Sistema	Número de conjuntos <i>fuzzy</i>	Função de pertinência
1	5	Triangular
2	5	Trapezoidal
3	5	Gaussiana
4	25	Triangular
5	25	Trapezoidal
6	25	Gaussiana
7	45	Triangular
8	45	Trapezoidal
9	45	Gaussiana

Fonte: Elaboração própria do autor.

### 3.5 Implementação computacional

A Tabela 3 apresenta informações dos pacotes utilizados para a implementação computacional no *software R* (R DEVELOPMENT CORE TEAM., 2020) modelo logito cumulativo com chances proporcionais (MLC), da rede neural artificial (RNA) e sistema baseado em regras *fuzzy* (SBRF).

Tabela 3 – Informações dos pacotes utilizados no *software R*.

	Pacote	
MLC	MASS	A função <i>polr</i> foi utilizada para construir o modelo de regressão ordinal.
RNA	neuralnet	A função <i>neuralnet</i> foi utilizada para construir as redes neurais artificiais. Foi utilizado o algoritmo de aprendizado RPROP– <i>Resilient Propagation</i> (RIEDMILLER; BRAUN, 1993).
SBRF	frbs	Foi utilizado o algoritmo de aprendizado WM (WANG; MENDEL, 1992).

Fonte: Elaboração própria do autor.

## 4 RESULTADOS E DISCUSSÃO

No tratamento inicial dos dados, foram retirados 28 *outliers*, resultando num conjunto de 188 observações. As variáveis dias de atraso, idade do cliente e número de parcelas contratadas foram agrupadas em faixas, como apresentadas na Tabela 4. A Tabela 5 apresenta um resumo das variáveis IGPM, valor de entrada e taxa de desemprego. A variável nível de risco do cliente apresentou 145 clientes com risco baixo (clientes com uma parcela em atraso), 21 clientes com risco moderado (clientes com duas parcelas em atraso) e 22 clientes com risco alto (clientes com mais de duas parcelas em atraso).

Tabela 4 – Variáveis agrupadas por faixas.

Dias de atraso		Idade do cliente		Número de parcelas contratadas	
Faixa	Intervalo	Faixa	Intervalo	Faixa	Intervalo
A	[0,30)	1	[21,35)	1	[0,50)
B	[30,60)	2	[35,49)	2	[50,100)
C	[60,90)	3	[49,67)	3	[100,155)
D	[90,150)				
E	[150,250)				
F	[250,500)				
G	[500,800)				

Fonte: Elaboração própria do autor.

Tabela 5 – Resumo das variáveis.

	IGPM (%)	Valor de entrada (R\$)	Taxa de desemprego (%)
Mínimo	-2,670	1000	6,50
Média	3,398	4843	11,51
Máximo	9,260	40000	13,70

Fonte: Elaboração própria do autor.

Considerando os dados de treinamento, o modelo logito cumulativo com chances proporcionais apresentou 87,61% de acerto, enquanto que, considerando os dados de teste, o modelo apresentou 81,33% de acerto.

A Tabela 6 apresenta as médias das porcentagens de acerto e média de erros referentes as diferentes arquiteturas de redes neurais propostas nas etapas de treinamento e teste, considerando 50 simulações em cada arquitetura.

Como critério de seleção da melhor arquitetura de rede neural, foi considerada a maior porcentagem de acerto na etapa de teste. A arquitetura 1 apresentou melhor desempenho, considerando uma camada escondida com 5 neurônios, obtendo médias iguais a 95,70% e 72,13% nas etapas de treinamento e teste, respectivamente. Considerando o algoritmo de aprendizado RPROP–*Resilient Propagation*, o processo de treinamento foi parado quando todas as derivadas

parciais absolutas da função de erro atingiram valores menores que 0,01 (valor padrão assumido pelo *software R*). Vale ressaltar que a menor média de erros foi obtida com a arquitetura 6, considerando 2 camadas escondidas com 25 e 10 neurônios, respectivamente.

Tabela 6 – Médias das porcentagens de acerto e média de erros nas etapas de treinamento e teste - redes neurais artificiais.

Arquitetura	Número de camadas escondidas	Número de neurônios nas camadas escondidas	Média Erro	Média Treinamento	Média Teste
1	1	5	3,53	95,70%	72,13%
2	1	15	3,32	95,68%	71,89%
3	1	25	3,18	95,72%	71,92%
4	2	5 - 2	4,12	93,82%	70,72%
5	2	15 - 5	3,31	95,86%	71,71%
6	2	25 - 10	3,05	95,89%	72,11%

Fonte: Elaboração própria do autor.

A Tabela 7 apresenta as médias das porcentagens de acerto nas etapas de treinamento e teste referentes aos sistemas baseados em regras *fuzzy*. O sistema 6, que considerou 25 conjuntos *fuzzy* com função de pertinência gaussiana, foi o que apresentou melhor desempenho, com médias das porcentagens de acerto iguais a 100% e 72%, referentes as etapas de treinamento e teste, respectivamente. Os sistemas com função de pertinência do tipo triangular apresentaram menores médias de porcentagens de acerto nas etapas de treinamento e teste.

Tabela 7 – Médias das porcentagens de acerto nas etapas de treinamento e teste - sistema baseado em regras *fuzzy*.

Sistema	Número de conjuntos <i>fuzzy</i>	Função de pertinência	Média Treinamento	Média Teste
1	5	Triangular	27,43%	34,67%
2	5	Trapezoidal	84,96%	48,00%
3	5	Gaussiana	84,96%	65,33%
4	25	Triangular	22,12%	61,33%
5	25	Trapezoidal	99,11%	54,67%
6	25	Gaussiana	100%	72,00%
7	45	Triangular	21,24%	65,33%
8	45	Trapezoidal	100%	64,00%
9	45	Gaussiana	100%	70,67%

Fonte: Elaboração própria do autor.

Os modelos propostos apresentaram resultados satisfatórios para a previsão de crédito de uma carteira de clientes de uma loteadora do sul de Minas Gerais. Para Picinini, Oliveira e Monteiro (2003) e Selau e Ribeiro (2009), porcentagens de acerto acima de 65% são consideradas boas para especialistas. Vale salientar que o modelo logito cumulativo com chances proporcionais apresentou resultado superior aos demais modelos na etapa de teste.

## 5 Considerações Finais

Devido ao aumento de inadimplência no Brasil e aos problemas que ela pode causar em empresas de médio e pequeno porte, este trabalho teve como objetivo apresentar três modelos para previsão de risco de crédito, baseados em regressão logística ordinal, redes neurais artificiais e sistemas baseados em regras *fuzzy*, em uma carteira de clientes de uma empresa de loteamentos de médio porte do sul de Minas Gerais.

O modelo de regressão logística ordinal obteve 81,33% de acertos na etapa de teste. Dentre as arquiteturas de redes neurais artificiais propostas, a rede de melhor desempenho foi realizada com uma camada escondida com 5 neurônios. A média de porcentagens de acertos foi igual a 72,13% na etapa de teste. Nos sistemas baseados em regras *fuzzy*, o melhor resultado foi obtido pelo sistema com 25 conjuntos *fuzzy* e função de pertinência gaussiana, com média de porcentagens de acertos igual a 72% na etapa de teste. Dentre os modelos considerados, a regressão logística ordinal apresentou melhor desempenho na etapa de teste.

O ponto limitante neste trabalho foi o número reduzido de variáveis. Por fim, esta pesquisa propõe para futuros trabalhos, utilizar um número maior de variáveis e até mesmo outros modelos de previsão de crédito para fins de comparação.

# Referências

- AGRESTI, A. *Analysis of Ordinal Categorical Data*. 2. ed. New Jersey: John Wiley & Sons, 2010.
- AGRESTI, A. *Categorical data analysis*. 3. ed. New Jersey: John Wiley & Sons, 2013.
- BARROS, L. C.; BASSANEZI, R. C. *Tópicos de lógica fuzzy e biomatemática*. 2. ed. Campinas: IMECC-UNICAMP, 2010.
- BEZERRA, P. *Inadimplência: a saga dos milhões de brasileiros negativados*. 2019. Acesso em: 19 de outubro de 2019. Disponível em: <<https://www.creditas.com.br/exponencial/inadimplencia-no-brasil/>>.
- BORBA, J. A.; DILL, R. Análise da rentabilidade de empresas: uma abordagem baseada na lógica nebulosa (fuzzy logic). *RAC - Eletrônica*, v. 1, n. 1, p. 47–66, 2007.
- BRITO, G. A. S.; Assaf Neto, A. Modelo de classificação de risco de crédito de empresas. *Revista Contabilidade & Finanças*, v. 19, n. 46, p. 18–29, 2008.
- CAMARGOS, M. A.; ARAÚJO, E.; CAMARGOS, A. M. C. S. A inadimplência em um programa de crédito de uma instituição financeira pública de Minas Gerais: uma análise utilizando regressão logística. *Revista de Gestão*, v. 19, n. 3, p. 473–492, 2012.
- CAMARGOS, M. A. et al. Fatores condicionantes de inadimplência em processos de concessão de crédito a micro e pequenas empresas do Estado de Minas Gerais. *Revista de Administração Contemporânea*, v. 14, n. 2, p. 333–352, 2010.
- CDNL. *Inadimplência do consumidor desacelera e cresce 0,9% no primeiro semestre de 2019, aponta indicador CNDL/SPC Brasil*. 2019. Acesso em: 23 setembro de 2019. Disponível em: <<http://site.cndl.org.br/>>.
- CORRAR, L.; PAULO, E.; FILHO, J. *Análise Multivariada para os cursos de Administração, Ciências Contábeis e Economia*. São Paulo: Atlas, 2009.
- FREITAS, A. A. C.; TAVARES, R.; MONTINI, A. A. Modelo inteligente para avaliação de risco de crédito pessoal baseado na lógica fuzzy. In: *XXXII Encontro da ANPAD*. Rio de Janeiro: [s.n.], 2008. p. 1–11.
- GONÇALVES, E. B.; GOUVÊA, M. A.; MANTOVANI, D. M. N. Análise de risco de crédito com o uso de regressão logística. *Revista Contemporânea de Contabilidade*, v. 10, n. 20, 2013.
- LUCENA, W. G. L. et al. Fatores que influenciam o endividamento e a inadimplência no setor imobiliário da cidade de Toritama-PE à luz das finanças comportamentais. *HOLOS*, v. 6, p. 90, 2015.
- MACHADO, M.; CERETTA, P.; VIEIRA, K. A Relação entre as Variáveis Macroeconômicas e a Concessão de Crédito no Mercado Imobiliário Brasileiro. *Revista de Gestão, Finanças e Contabilidade*, v. 4, n. 3, p. 64–84, 2014.

- MENDONÇA, M. J. C. de. O crédito imobiliário no Brasil e sua relação com a política monetária. *Revista Brasileira de Economia*, v. 67, n. 4, p. 457–495, 2013.
- NELDER, J.; WEDDERBURN, R. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, v. 135, n. 3, p. 370–384, 1972.
- PALMUTI, C. S.; PICCHIAI, D. Mensuração do risco de crédito por meio de análise estatística multivariada. *Revista Economia Ensaios*, v. 26, n. 2, p. 7–22, 2012.
- PICININI, R.; OLIVEIRA, G. M. B.; MONTEIRO, L. H. A. Mineração de critério de credit scoring utilizando algoritmos genéticos. In: *Simpósio Brasileiro de Automação Inteligente: Anais...* Bauru: Universidade de Brasília, 2003. v. 6.
- R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing*. 2020. Acesso em: 10 de março de 2020. Disponível em: <<http://www.r-project.org>>.
- RIEDMILLER, M.; BRAUN, H. A direct adaptive method for faster backpropagation learning: The rprop algorithm. *Proceedings of the IEEE International Conference on Neural Network*, p. 586–591, 1993.
- ROCHA, R. *Número de inadimplentes no Brasil é maior do que a população de quase todos os países da América Latina*. 2019. Acesso em: 17 de novembro de 2019. Disponível em: <<https://extra.globo.com/noticias/economia/>>.
- SCHRICKEL, W. K. *Análise de crédito: concessão e gerência de empréstimos*. São Paulo: Atlas, 1995.
- SELAU, L. P. R.; RIBEIRO, J. L. D. Uma sistemática para construção e escolha de modelos de previsão de risco de crédito. *Gestão & Produção*, v. 16, n. 3, p. 398–413, 2009.
- SILVA, I.; SPATTI, D.; FLAUZINO, R. *Redes Neurais Artificiais para engenharia e ciências aplicadas*. São Paulo: Artliber, 2010.
- STEINER, M. T. A. et al. Extração de regras de classificação a partir de redes neurais para auxílio à tomada de decisão na concessão de crédito bancário. *Pesquisa Operacional*, v. 27, n. 3, p. 407–426, 2007.
- SYAU, Y.; HSIEH, H.; LEE, E. S. Fuzzy number in the credit rating of enterprise financial condition. *Review of Quantitative Finance and Accounting*, v. 17, n. 4, p. 351–360, 2001.
- TANAKA, K. *An introduction to fuzzy logic for practical applications*. Tokyo: Springer, 1997.
- VALENTE, G. F. S. et al. Aplicação de redes neurais artificiais como teste de detecção de fraude de leite por adição de soro de queijo. *Revista do Instituto de Laticínios Cândido Tostes*, v. 69, n. 6, p. 425, 2014.
- WANG, L. X.; MENDEL, J. M. Generating fuzzy rule by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 22, n. 6, p. 1414–1427, 1992.
- WUERGES, A. F. E.; BORBA, J. A. Redes neurais, lógica nebulosa e algoritmos genéticos: aplicações e possibilidades em finanças e contabilidade. *JISTEM Journal of Information Systems and Technology Management*, v. 7, n. 1, p. 163–182, 2010.
- ZADEH, L. Fuzzy sets. *Journal Information and Control*, n. 8, p. 338–353, 1965.