

Proposição de modelos de previsão de risco de crédito para pequenas e médias empresas por meio da Regressão Logística

Proposition of credit risk forecasting models for small and medium enterprises through logistic regression

Flávio Führ¹ⁱ

Orcid: <https://orcid.org/0000-0003-3205-4583>

José Donizetti de Lima²ⁱⁱ

Orcid: <https://orcid.org/0000-0001-5260-9035>

Gilson Ditzel Santos³ⁱⁱⁱ

Orcid: <https://orcid.org/0000-0002-7839-2281>

Sady Mazzioni^{4iv}

Orcid: <https://orcid.org/0000-0002-8976-6699>

Resumo

A busca por padrões que contribuam na predição de risco, é crescente nas organizações. A utilização de modelos de *credit scoring* busca auxiliar o analista de crédito na tomada de decisão. Este trabalho objetiva elaborar procedimentos metodológicos, para estruturar e melhorar os modelos de *credit scoring* direcionados a análise de pequenas e médias empresas. Com a utilização da técnica estatística da regressão logística, por meio das melhorias elaboradas nos procedimentos metodológicos, como exemplo: divisão da base de dados em classes conforme enquadramento das empresas, foi possível o desenvolvimento de 5 modelos de *credit scoring*, sendo um modelo para cada classe de empresas e outro para a base geral de dados. Os modelos foram direcionados às entidades de fomento e concessão de crédito para pequenas e médias empresas. As acurácias dos modelos apresentaram percentuais expressivos para base de dados com variáveis não contábeis e não auditáveis, atingindo percentuais satisfatórios.

Palavras-chave: *credit scoring*; pequenas e médias empresas; regressão logística.

Abstract

The search for standards that contribute to the prediction of risk is growing in organizations. The use of credit scoring models seeks to assist the credit analyst in making decisions. This work aims to develop methodological procedures, to structure and improve credit scoring models aimed at the analysis of small and medium-sized companies. With the use of the statistical technique of logistic regression, through the improvements developed in the methodological procedures, such as division of the database into classes according to the companies' framework, it was possible to develop 5 credit scoring models, one model for each class of companies and another for the general database. The models were directed to entities that promote and grant credit to small and medium-sized companies. The accuracy of the models showed significant percentages for the database with non-accounting and non-auditable variables, reaching satisfactory percentages.

Keywords: credit scoring; small and medium companies; logistic regression.

¹ Instituto Federal do Paraná. Colegiado de Ciências Contábeis, Brasil. E-mail: flavio.fuhr@ifpr.edu.br

² Universidade Tecnológica Federal do Paraná – UTFPR, Campus Pato Branco, Brasil. E-mail: donizetti@utfpr.edu.br

³ Universidade Tecnológica Federal do Paraná – UTFPR, Campus Pato Branco, Brasil. E-mail: ditzel@utfpr.edu.br

⁴ Universidade Comunitária da Região de Chapecó - Unochapecó. Área de Ciências Sociais Aplicadas, Brasil. E-mail: sady@unochapeco.edu.br

1 INTRODUÇÃO

Fatores como o aumento do grau de estabilidade econômica, surgimento de novos produtos e serviços e controle da inflação, contribuem para a ampliação do mercado consumidor (VENTURA, 2010). Esta inclusão de pessoas e empresas, no mercado nacional brasileiro, reposiciona a análise de crédito e seu grau de importância, uma vez que as empresas optam por comercializar seus produtos e serviços a prazo, necessitando de critérios de avaliação no momento da concessão, pois o risco de inadimplência está incluso no processo de venda.

A avaliação do risco tem por objetivo melhorar a qualidade de uma carteira de clientes, favorecendo uma venda saudável e evitando ao máximo a perda de valores, sobre créditos fornecidos de forma equivocada ou a clientes que geram prejuízos aos negócios. Empresas que possuem boa avaliação, levam vantagens sobre seus concorrentes (GOUVÊA; GONÇALVES, 2007).

Uma das principais bases informacionais é a contabilidade, a qual deve possuir confiabilidade das informações apresentadas e ser relevante para que o resultado da análise de concessão seja útil. Argumenta Bruni (2010) que muitas das informações trabalhadas no setor financeiro são registradas e armazenadas na contabilidade.

No Brasil, as pequenas empresas, utilizam a estrutura contábil para cumprimento das obrigações fiscais, gerando poucas informações para o setor de gerenciamento e decisão da organização. Ressalta Berti (2012) que a análise por meio das demonstrações contábeis, das micro e pequenas empresas, apresenta o inconveniente das estruturas contábeis não conseguirem refletir a realidade das transações, tendo em vista que não registra integralmente as informações, utilizando a contabilidade apenas para cumprir exigências fiscais.

Historicamente, as pequenas empresas enfrentam dificuldades em acessar financiamento devido à falta de credibilidade nas informações, sendo que por vezes não possuem demonstrações contábeis auditadas e certificadas em base regular (BERGER; COWAN; FRAME, 2011).

As instituições que trabalham focadas nas pequenas e médias empresas (PMEs), compartilham do dilema descrito por Berger, Cowan e Frame (2011) e Berti (2012), referente à credibilidade e auditoria das informações contábeis, utilizando de outras fontes informacionais para a análise de crédito (variáveis). A modelagem de crédito apresenta-se como um instrumento de apoio aos analistas, porém, há poucos estudos sobre modelagem de crédito para empresas de pequeno e médio porte, destacando a importância dessas empresas nas economias dos países ao redor do mundo (LI; NISKANEN; KOLEHMAINEN, 2016). Adicionalmente, há que se considerar a inexistência de um método unânime ou global para lidar com problemas de pontuação de crédito, havendo crescente interesse pelas organizações no uso de conjuntos de classificação (MARQUÉS; GARCÍA; SÁNCHEZ, 2012).

A partir do contexto apresentado, pergunta-se: Como avaliar a concessão do crédito à PME a partir da avaliação de variáveis não contábeis?

Com o intuito de contribuir com o setor de análise de crédito, o objetivo do estudo é elaborar modelos de *credit scoring*, para avaliação de PME com variáveis não contábeis. Os objetivos específicos são: (i) desenvolver procedimentos metodológicos para manipulação das variáveis conforme enquadramento das empresas em Microempreendedor Individual (MEI), Microempresa (ME), Pequena Empresa (PE), Médias Empresas (MédE). (ii) desenvolver os modelos de *credit scoring*, para as 4 bases de dados: MEI, ME, PE e MédE, por meio da técnica da Regressão

Logística, e; (iii) analisar a potencialidade dos modelos, como fonte de informação manipulada, voltada a auxiliar a tomada de decisões.

A originalidade do trabalho encontra-se em três aspectos: Primeiro, utilizando-se base de informações secundárias, não contábil, segregou-se a amostra por faixa de faturamento utilizando parâmetros do Banco Nacional de Desenvolvimento Econômico e Social (BNDES 2010) e Lei da Microempresa nº 128/2008, melhorando a caracterização das empresas dentro de seu âmbito de atuação. Segundo, apresenta quatro modelos de *credit scoring*, sendo um para cada classe de empresas, no qual observou-se que três deles tem boa aderência ao auxílio decisório, com predição acima de 80% na probabilidade da identificação do cliente em inadimplente ou adimplente. Terceiro, foi possível identificar as variáveis que influenciaram os modelos, oportunizando a instituição voltar atenções a estas quando do colhimento de informações cadastrais ou de crédito.

A contribuição teórica do trabalho, encontra-se no desenvolvimento de um modelo de *credit scoring*, para PMEs com dados não contábeis, mais próximo de sua realidade. Na prática, busca-se contribuir com análise das instituições que possuam portfólio de PME em suas carteiras, fomentando-as com volume de indicadores no suporte a tomada de decisão.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 *Rating versus Credit Scoring*

As informações, sejam cadastrais, creditícias ou contábeis, são fatores relevantes nos modelos de classificação “*rating*”, palavra bastante utilizada na montagem matemática e estatística para enquadramento de crédito. Destaca Silva (2004), que *rating* é uma avaliação da informação feita por meio da mensuração e ponderação de variáveis determinantes, fornecendo uma graduação. Relatam

Rogers, Dany; Mendes-da-Silva e Rogers, Pablo (2016), que apesar de vários estudos que analisam a relação entre classificação de crédito e estrutura de capital, ainda é pouco estudado nos ambientes institucionais da América Latina [...], já *Credit Scoring* objetiva avaliar o risco de inadimplência com base numa pontuação relacionada à probabilidade de um candidato cair na classe ruim (KELLY; HAND, 1999), mais especificamente no Brasil, a utilização de *credit scoring*, tiveram maior interesse dos pesquisadores apenas nos últimos anos (CAMARGOS, Marcos; CAMARGOS, Mirela; ARAÚJO, 2012).

A classificação de crédito busca criar um modelo que transcreva informações quantitativas e qualitativas da credibilidade da empresa refletir a qualidade de um devedor (MILERIS, 2012). Para o desenvolvimento da modelagem, a literatura indica a utilização de técnicas, a exemplo da Análise Envoltória de Dados (DEA), Análise Discriminante (AD), Regressão Logística (RL) ou Modelo do Regressão Logístico Múltiplo (MRLM), Redes Neurais Artificiais (RNA), Árvore de Decisão (DT), Máquina de Suporte Vetor (SVM), e Grupo de Manipulação de Dados (GMDH).

2.2 Regressão Logística

Um dos pioneiros a trabalhar a técnica RL para modelagem de previsão de risco, foi James A. Ohlson, em 1980, identificando sua potencialidade e facilidade de aplicação. A RL é uma das técnicas estatísticas mais utilizadas para elaboração de modelo de *credit scoring*, no qual a variável dependente é binária (LUO; WU, Desheng; WU, Dexiang, 2016). Em uma perspectiva prática, a RL é de fácil entendimento, com parâmetros simples para sua implantação, especialmente vantajosa para a formulação da pontuação preditiva (LUO; KONG; NIE, 2016).

Conforme argumentam Dias Filho e Corrar (2017), a equação da RL, calcula a

probabilidade relativa à ocorrência de determinado evento ou a “probabilidade associada a cada observação em razão de chance (*odds ratio*), que representa a

probabilidade de sucesso comparada com a de fracasso”, expressa da forma apresentada na equação:

$$\text{Razão de chance} = \frac{P(\text{sucesso})}{1 - P(\text{sucesso})} \quad (1)$$

Ainda, para Dias Filho e Corrar (2017), de forma mais simplificada, a

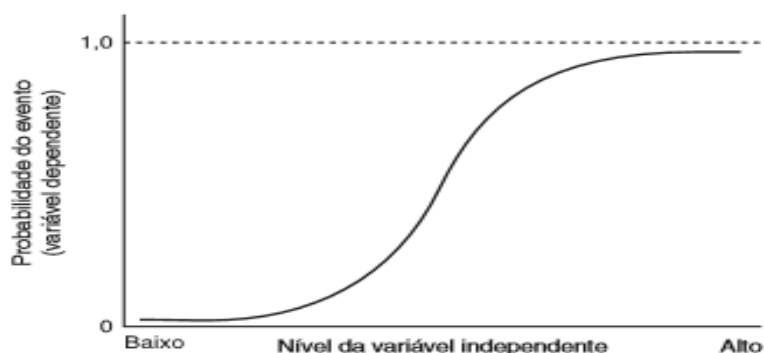
equação logística pode assumir o formato apresentado na equação:

$$P(\text{evento}) = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_{ki})}} \quad (2)$$

A RL utiliza a curva logística para demonstrar a relação entre as variáveis independentes e dependentes utilizando o intervalo entre 0 e 1 (HAIR JR. *et al.*, 2009). Essa probabilidade de ocorrência ou previsão, está baseada nos valores das variáveis independentes e nos coeficientes

estimados. Pode-se caracterizar que se a probabilidade prevista for maior que 0,50, então, tem-se uma previsão de que o resultado seja 1 (o evento ocorreu); caso contrário, o resultado é previsto como sendo 0 (o evento não ocorreu) (Hair JR. *et al.*, 2009). A Figura 1 ilustra a RL.

Figura 1 – Forma de relação logística entre variáveis dependentes e independentes



Fonte: (HAIR JR. *et al.*, 2009, p. 285).

2.3 Trabalhos de modelagem de *credit scoring*

Os trabalhos de modelagens de crédito utilizam várias técnicas, entre elas a RL, sobre dados elaborados das empresas a partir das informações geradas pelas demonstrações contábeis. Esses dados são retirados de empresas que possuem uma contabilidade estruturada de informações, capazes de gerar indicadores financeiros e contábeis.

Esses indicadores geram informações que posteriormente são tratadas como variáveis na aplicação de um modelo de *credit scoring*. Para as empresas que não possuem obrigatoriedade das publicações da informação, os modelos de pontuação de crédito utilizam dados oriundos da ficha cadastral do cliente, dados de crédito e histórico interno. O Quadro 2 apresenta trabalhos desenvolvidos na área de modelagem, voltadas à modelagem de *credit Scoring*.

Quadro 2 - 0 Trabalhos desenvolvidos na área de modelagem de *credit Scoring*

Autor	Objetivo	Variáveis	Técnica	Acurácia %	Fonte
Li, Niskanen e Kolehmainen (2016)	Elaborar modelo híbrido RN e RL	Rentabilidade, retorno, liquidez e da atividade	RL	71,2	PME's China
			RNA	79,2	
			Híbrido	83,1	
Zhu <i>et al.</i> (2016)	Prever de risco de crédito para pequenas empresas	Índices Financeiros e não financeiros,	RL	61,3	PME's China
			RNA	68,8	
			Híbrido I	70,2	
			Híbrido II	88,5	
Smaranda C. (2014) **	Testar os modelos de previsão de falências	Indicadores financeiros	RL	87,2	PME's Europa Central/Oriental
			AD	83,0	
Ciampi F.; Gordini N. * (2013)	Testar um modelo de previsão padrão para PMEs	Variáveis contábeis	AD	68,8	PME's Italianas
			RL	68,5	
			RNA	72,8	
Selau; Ribeiro (2011)	Propor uma abordagem sistemática para construir modelo de revisão de risco de crédito	Dados de cadastro	AD	73,2	Rede de Farmácias
			RL	73,3	
			RNA	74,8	
Wang <i>et al.</i> (2011)	Verificar a eficiência tradicional das variáveis financeiras para prever a inadimplência das PMEs	Financeiras e qualitativas	RL Model II	88,6	PME's China
			RL Model III	100	
Marques e Lima (2002)	Classificar informações adimplentes ou inadimplentes	Econômicas e financeiras	AD	91,6	BRDE Brasil
			RL	93,1	

Fonte: Dados do Estudo

Nota. * Os percentuais de acurácia foram extraídos da Tabela 17 do trabalho de Ciampi; Gordini (2013, p.40),

Nota** Foram extraídos apenas 2 percentuais de acurácia, conforme Tabela 1 e Quadro 3 do trabalho de Smaranda (2014).

Torna-se importante observar o grau de significância dos resultados alcançados pelos modelos apresentados no Quadro 2, havendo variação entre 61,3% e 93,1% de acertos sobre as técnicas utilizadas pelos autores, (AD, RL, RNA e suas variações HÍBRIDAS). Os trabalhos que utilizaram a técnica de RL, obtiveram resultados expressivos, demonstrando eficiência na identificação de variáveis.

Observa-se que dos trabalhos relacionados no Quadro 2, apenas o trabalho de Selau e Ribeiro (2011) utiliza dados de cadastro (voltado para analisar crédito de pessoa física), os demais utilizaram, no todo ou em parte, informações oriundas da contabilidade. Essa visão reforça a importância do estudo para concessão de crédito direcionada à pessoa jurídica com utilização de informações não contábeis, devido à dificuldade de obtenção da informação ou falta dela para PME.

3 Metodologia

Esta pesquisa é de natureza aplicada, a qual busca a solução imediata de problemas concretos do cotidiano por meio da orientação prática visando à gestão e tomada de decisão (DUARTE; FURTADO, 2014). Quanto aos objetivos de pesquisa, é caracterizada como explicativa. Segundo Gil (2009), na busca de responder o “porquê” das coisas, a pesquisa explicativa visa identificar fatores que contribuem para a ocorrência dos fenômenos e conhecimento da realidade exigindo descrição suficiente e detalhada.

3.1 Procedimentos Metodológicos

Quanto aos procedimentos, para desenvolvimento dos modelos de *credit scoring*, foram aplicados 11 passos, os quais são apresentados no Quadro 3 e explicados na sequência.

Quadro 3 – 11 procedimentos metodológicos

Passo 1	Análise da base de dados
Passo 2	Eliminação dos <i>missings values</i>
Passo 3	Classificação das Variáveis
Passo 4	Classificação das Empresas em Categorias (4 classes)
Passo 5	Discretização
Passo 6	Transformação de variáveis em <i>dummy</i>
Passo 7	Aplicação do método <i>Stepwise</i>
Passo 8	Separação da base em: bases teste base principal
Passo 9	Pressupostos da técnica RL
Passo 10	Elaboração dos Modelos com a técnica RL
Passo 11	Acurácia da técnica RL

Fonte: Dados do Estudo

Passo 1: Análise da base de dados - As observações e dados foram coletados por meio de relatórios disponíveis na instituição financeira, vocacionada para o crédito comercial e industrial. O nome e outras informações de identificação dos clientes foram substituídos por numerações aleatórias para garantir o anonimato.

Passo 2: Eliminação de dados ausentes (*missings values*) - Nessa etapa foi efetuada uma análise do banco de dados, sendo eliminados 219 contratos da totalidade de 1.710, restando uma base de 1.491 contratos. A análise focou-se na busca de informações faltantes nas variáveis independentes para todos os contratos. O período de levantamento dos dados corresponde aos créditos ativos até 10 de outubro de 2017. Esse processo de eliminação de dados ausentes também foi adotado por (CAMARGOS, Marcos; CAMARGOS, Mirela; ARAÚJO, 2012).

O processo de “dados ausentes ou perdidos”, para Rodrigues e Paulo (2017), corresponde a qualquer evento sistemático externo ao respondente, levando primariamente o pesquisador a buscar as razões inerentes a esses. A justificativa para eliminação dos dados, encontra-se na necessidade da informação para classificação do porte das empresas. Destaca-se que os dados ausentes não foram captados devido a fazer parte de outro sistema (base de dados), não sendo possível acessar na data do levantamento dos dados.

Passo 3: Classificação das Variáveis – A variável resposta refere-se à qualidade do crédito, sendo identificado como adimplente ou inadimplente. Conforme política da instituição, são considerados inadimplentes, aqueles que apresentaram atraso superior a 30 dias ininterruptos dentro do ano fiscal. Já, os clientes identificados com pagamentos sem atrasos ou com atrasos iguais ou inferiores a 30 dias, foram classificados como adimplentes.

Preparação dos dados: As variáveis quantitativas podem ser medidas em escalas apresentadas da forma discreta ou contínua. As variáveis qualitativas não possuem valores quantitativos, também conhecidas como categóricas, podendo ser classificadas em nominal e ordinal (RODRIGUES; PAULO, 2017). Devido à peculiaridade da base de dados referir-se a uma instituição específica cooperativa e creditícia, algumas variáveis são autoexplicativas, porém outras necessitam de esclarecimento, são elas: Código do município – identifica a unidade da federação; Cnae – é o código que identifica a atividade econômica, também pode ser utilizado para facilitar o enquadramento da empresa; Valor da Garantia – especifica o valor que o cliente tem em garantia das operações de crédito; Risco – é o enquadramento da empresa junto à instituição financeira da probabilidade de inadimplência; N° de Produtos – especifica a quantidade de produtos que o cliente adquiriu da

instituição financeira, em vigor na data do levantamento de dados; Valor da Cota Capital – corresponde ao valor, em moeda nacional, que o cliente tem depositado por ocasião da aquisição das cotas da cooperativa; Porte das Empresas – identifica o enquadramento da empresa,

mediante a renda/faturamento apresentada pela empresa que encontra-se em análise.

O Quadro 4 apresenta as variáveis que foram utilizadas nos modelos, suas ordens e codificações oriundas da base de dados fornecidas pela Instituição Financeira, denominadas de variáveis primitivas.

Quadro 4 - Variáveis primitivas extraídas da base de dados

Código	Variável	Ordem	Código	Variável	Ordem
V1	Código Município	Nominal	V9	Nº de Produtos	Contínua
V2	Cnae	Nominal	V10	Valor Cota Capital	Contínua
V3	Renda Mensal	Contínua	V11	Idade	Ordinal
V4	Valor das Garantia	Discreta	V12	Renda Anual	Contínua
V5	Risco	Ordinal	V13	Tempo de Filiação	Discreta
V6	Valor do Contrato	Contínua	V14	Saldo Devedor	Contínua
V7	Nº Parcela	Contínua	V15	Porte das Empresas	Nominal
V8	Dias Atraso	Contínua			

Fonte: Dados do Estudo

Parte das variáveis utilizadas neste estudo também foram encontradas nos estudos de

outros autores, conforme demonstrado no Quadro 05.

Quadro 05 – Variáveis deste estudo, identificadas em outros trabalhos

Autores	Variáveis identificadas em outros trabalhos	Variáveis deste estudo	Observações
Silva, Ribeiro e Matias (2016)	Montante do crédito requerido, duração do empréstimo requerido, idade, entre outras	(V6) (V7) (V11)	Foco na pessoa física
Gonçalves, Gouvêa e Mantovani (2013)	Idade, salário (compatível com renda), valor do empréstimo, número de parcelas, entre outras	(V11) (V12) (V16)	As variáveis quantitativas foram transformadas em variáveis categóricas
Camargos Marcos, Camargos Mirela e Araújo (2012)	Setor da atividade, valor dos bens do avalista, valor do financiamento, tempo de atividade da empresa, valor do faturamento anual da empresa, entre outras	(V2) (V4) (V6) (V12)	Voltado para Pessoa Jurídica
Araújo e Carmona (2009)	Natureza da atividade econômica do negócio, número de parcelas do último empréstimo, idade, receita bruta do negócio, tempo de funcionamento, entre outras	(V2) (V7) (V11) (V12)	Variáveis qualitativas foram inseridas por meio das variáveis dummy
Lemos; Steiner; Nievola (2005)	Setor de atividade, risco atribuído, tempo de atividade, faturamento bruto anual, entre outras	(V2) (V5, (V11) (V12)	Voltado para Pessoa Jurídica

Fonte: Dados do Estudo

As variáveis (V9) e (V10), não foram encontradas nos trabalhos relacionados à Pessoa Jurídica constante do Quadro 5.

Passo 4: Divisão das classes de empresas
 - A importância de reconhecer as diferenças existentes entre as classes de empresas

possibilita analisá-las, mais especificamente, dentro de suas formas de atuação e gestão. Também entendendo que, quando em situações de recessão de mercado, geralmente são as primeiras a apresentar dificuldades e as que levam um

tempo maior para recuperação (SILVA, 2016).

As Pequenas Empresas, caracterizadas como: Microempreendedor Individual (MEI), Microempresa (ME), Pequena Empresa (PE), diferem das Médias Empresas (MédE), assim como diferem entre si, nas suas formas estruturais, barreiras, benefícios fiscais entre outros, apresentando realidades diferenciadas, gerando informações próprias dentro do seu universo de atuação. Nesse contexto, a segmentação em classes é relevante, pois possibilita tratar as variáveis e seus valores dentro da realidade de atuação das empresas no mercado.

Devido às peculiaridades da fonte de informação, próprias do formato de cada empresa, conforme descreve Alvim (1998), quanto ao desafio de disponibilizar informação adequada que possa subsidiar o processo da tomada de decisão; os modelos propostos foram elaborados a partir das características individuais de cada categoria de empresas, sendo: MEI, ME, PE e MédE. As empresas estão distribuídas conforme faixas (ou classes) do BNDES, (2010), na qual a faixa da MEI, foi incluída conforme Lei 128/2008, que alterou a Lei 123/2006, Lei da Microempresa. A classificação adotada é ilustrada no Quadro 6.

Quadro 6 - Enquadramento das empresas conforme renda anual

Categoria	MEI	ME	PE	MédE
Faixas	Até R\$ 60 mil	De R\$ 60 mil a R\$ 2,4 milhões	De R\$ 2,4 milhões a R\$ 16 milhões	De 16 milhos a R\$ 90 milhões

Fonte: Adaptado de BNDES (2010); Lei 128/2008 e Lei 123/2006

Para a modelagem da *credit scoring*, com a técnica de RL, utilizou-se o Software SPSS Statistics®. Foram elaborados 5 modelos de *credit scoring*, sendo um para cada classe de empresas e outro modelo para a base geral de dados, descrita como DG.

Passo 5: Discretização de variáveis - As variáveis idade, renda anual, tempo de filiação, saldo devedor, número de parcelas e valor do contrato, foram discretizadas. Argumentam, García *et al.* (2013), que a discretização pode ser observada como um método de redução de dados, reduzindo em subconjuntos um grande volume de dados, pois mapeia os dados de um enorme conjunto de valores numéricos para um subconjunto de valores discretos bastante reduzido. Como ponto de corte na média e desvio/padrão calculados foi (+/- 1 desvio/padrão) para criação dos subgrupos de dados dentro de cada categoria de empresas, ou seja: MEI, ME, PE, MédE e DG. Segundo Lunet, Severo e Barros (2006), o valor do desvio padrão reflete a variabilidade das observações em relação à

média, caracterizado como uma medida de dispersão.

Passo 6: Variáveis Dummies – Posteriormente à discretização, as variáveis: idade, renda anual, tempo de filiação, saldo devedor, número de parcelas e valor do contrato, foram transformadas em variáveis *dummy*, surgindo uma nova variável para cada faixa/classe discretizada, o que ocasionou o surgimento de 23 novas variáveis, em substituição às variáveis primitivas. Segundo Missio e Jacobi (2007), a variável *dummy* é uma variável artificial a qual assume valor igual a 0 ou 1, indicando a ausência ou presença de algum atributo, transformando o modelo de regressão em uma ferramenta flexível para lidar com problemas encontrados em estudos empíricos. Cunha e Coelho (2017), indicam que o uso da *dummy* pode melhorar o percentual do coeficiente de determinação (R²) e sua contribuição é indicar a presença ou ausência de determinado atributo, assumindo apenas 0 ou 1.

Passo 7: Seleção de variáveis – método *stepwise* - Para seleção das variáveis, foi

aplicado o método *stepwise*. Esse método estatístico permite determinar um conjunto de variáveis significativas, implicando na inclusão ou remoção de variáveis potenciais (DINIZ; LOUZADA, 2013). O *stepwise* é utilizado em métodos de estimação, com seleção sequencial de variáveis, objetivando identificar a variável independente com o maior poder preditivo no modelo de RL (HAIR JR. *et al.*, 2009).

Passo 8: Validação e ajuste dos modelos - Para as técnicas selecionadas, foram utilizadas uma base de teste (BT) de 20% dos dados totais, escolhidos aleatoriamente,

para validação dos modelos. Os 80% restantes da base, denominadas de base principal (BP), foram utilizados para ajuste e elaboração dos modelos. Esse procedimento foi efetuado para todas as 4 classes de empresas e para os DG. Esse percentual também é encontrado nos trabalhos de SELAU e RIBEIRO, 2009; SILVA; RIBEIRO; MATIAS, 2016.

Passo 9: Pressupostos da técnica RL - Na Tabela 1, são apresentados os pressupostos necessários para o desenvolvimento da técnica.

Tabela 1 - Pressupostos para utilização da técnica

Técnica	Termo	Pressupostos: (Dais Filho & Corrar, 2017)	Testes
Regressão Logística	Não Linear	a. Teste de correlação b. Ausência de multicolinearidade entre as variáveis independentes	a. Tolerância b. Inverso da Tolerância (VIF) c. Verossimilhança d. Hosmer e Lemeshow

Fonte: Elaborado pelos autores

Passo 10: Elaboração dos modelos - Com base na seleção das variáveis, com maior poder preditivo, após utilização do método *stepwise* e a técnica da RL, foram elaborados 5 modelos de *credit scoring*, um para cada classe de empresas: MEI, ME, PE, MédE e DG.

Passo 11: Acurácia da Técnica RL - Uma das formas utilizadas para verificação da capacidade dos modelos é por meio da sua capacidade de predição, também entendido como acurácia. Dentro do processo da RL é gerada a classificação da acurácia geral, sendo na sequência efetuada uma análise comparativa da acurácia (predição) geral da técnica da RL para as 4 classes de empresas *versus* DG.

4 resultados

A técnica RL foi aplicada sobre a base de dados de contratos de uma instituição financeira com corte temporal em 10 de outubro de 2017. A base de dados, inicialmente, foi separada por faixa de

faturamento, conforme ilustrado no Quadro 5. Ainda, com as informações coletadas, foi utilizada a base de dados para formulação de um modelo geral (DG).

4.1 Análise Preliminar

Inicialmente, antes da abordagem com as amostras de teste, coube uma observação mais detalhada do banco de dados, a fim de eliminar possíveis inconsistências decorrentes de *missing*, obtendo um volume total de 1.448 clientes, após a eliminação dos dados inconsistentes.

Para a variável dependente foi atribuído o valor de “0” aos Clientes/Empresas Inadimplentes e o valor de “1” para Clientes/Empresas Adimplentes. O ponto de corte utilizado para desenvolvimento da técnica RL foi 0,50, conforme destacado no item 2.3.

4.2 Análise das variáveis

Ao iniciar a análise da base de dados, foi observado o comportamento de

cada variável independente em relação à variável dependente e conseqüentemente entre as próprias variáveis independentes. As variáveis V3 (Renda Mensal) e V15 (Classificação), foram retiradas do modelo devido a serem inter-relacionadas com outras variáveis, apresentando correlação. A existência de correlação é determinada quando as variáveis independentes, entre si, explicam o mesmo fato com informações similares, este fenômeno é conhecido como multicolinearidade. Para Hair Jr. *et al.* (2009), a multicolinearidade, medida de tolerância, denota que duas ou mais variáveis independentes estão altamente correlacionadas, quando uma variável pode ser prevista por outras variáveis com baixo poder explicativo para o conjunto. Por sua vez, a variável V5 (Risco) é calculada, pela Instituição Financeira, com base na evolução da variável V8 (Dias de Atraso), sendo ambas descartadas.

4.3 Discretização das variáveis

Para um melhor ajuste dos modelos, observadas as classes das empresas, as variáveis, idade, renda anual, tempo de filiação, saldo devedor, número de parcelas e valor do contrato, foram discretizadas. Cada uma dessas variáveis, por meio da discretização, utilizando como parâmetro (+/-) 1 desvio/padrão, foi dividida em subconjuntos, dentro da mesma variável. Posteriormente, foi subdividido o grupo de dados gerais em grupos menores de 3 a 4 subconjuntos por variável.

4.4 Transformação das variáveis em *dummies*

A incorporação de variáveis *dummy*, aos modelos de regressão linear, os torna capaz de lidar com muitos problemas

encontrados de forma extremamente flexível, principalmente em estudos empíricos (MISSIO; JACOBI, 2007). As variáveis idade (V11), renda anual (V12), tempo de filiação (V13), saldo devedor (V14), número de parcelas (V7) e valor do contrato (V6), foram transformadas em variáveis *dummies*.

Para essa transformação, utilizaram-se os subconjuntos, já determinados pela discretização, como referência para as faixas de classificação, para cada nova variável artificial. Mediante este processo de transformação, houve o surgimento de 23 novas variáveis *dummy* em substituição às variáveis primitivas, V6, V7, V11, V12, V13 e V14, para cada classe de empresas, separadas por categoria de faturamento e para os DG.

4.5 Regressão Logística

A técnica de RL, utilizando uma variável dicotômica como variável dependente, tem sido bastante utilizada como instrumento de predição. Essa técnica possibilita contornar problemas como a homogeneidade de variância e a normalidade na distribuição dos erros (DIAS FILHO; CORRAR, 2017).

Para estruturação dos modelos de pontuação de crédito, utilizou-se de 5 variáveis primitivas sendo: (V1, V2, V4, V9 V10) e 23 variáveis *dummies*, também conhecidas como variáveis artificiais, em substituição às variáveis primitivas (V6, V7, V11, V12, V13 e V14), totalizando 28 variáveis, conforme descrito no Quadro 7. Para interpretação do Quadro 7, são destacadas as seguintes abreviaturas: P = Primitivas; MV = *Missing Values*; PS = Primitivas Substituídas; PU = Primitivas Utilizadas e Ar = Artificiais.

Quadro 7 - Descrição das variáveis utilizadas nos testes de RL

Variáveis															Cálculo		
V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	(=)	15	P
		V3		V5			V8							V15	(-)	4	MV
V1	V2		V4		V6	V7		V9	V10	V11	V12	V13	V14		(=)	11	P
					V6	V7				V11	V12	V13	V14		(-)	6	PS
V1	V2		V4					V9	V10						(=)	5	PU
V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27				(+)	23	Ar
V28	V29	V30	V31	V32	V33	V34	V35	V36	V37	V38					(=)	28	P e Ar
Total de Variáveis															(=)	28	P e Ar

Fonte: Dados do Estudo

Para seleção das variáveis independentes, as quais possuem o maior poder preditivo, a fim de incorporar o modelo de previsão para cada classe de empresas, utilizou-se o método *stepwise*, conforme opção disponibilizada pelo Software SPSS®. O nível de corte adotado foi de 0,50 para a significância da seleção e agrupamento de contratos “cliente”, onde o adimplente assumiu a codificação de valor “0” e o inadimplente o valor “1”, no volume

das 28 variáveis independentes utilizadas como entrada nos modelos de pontuação de crédito, após utilização do método *stepwise*.

Conforme já citado, esse método efetua uma seleção sequencial, objetivando identificar a variável que apresenta um maior poder preditivo para a regressão (HAIR JR. *et al.*, 2009). Obteve-se uma saída de 9 variáveis com poder de discriminação, conforme apresentado na Tabela 2.

Tabela 2 - Variáveis inclusas no modelo de RL utilizando o método *stepwise*

Variáveis	MEI	ME	PE	MédE	DG
V2	-	-	-	-	X
V9	X	X	X	X	X
V19	X	X	-	X	X
V21	-	X	-	-	-
V26	-	-	-	-	X
V27	-	X	-	-	-
V34	-	-	-	X	-
V35	-	X	-	-	-
V36	-	-	-	-	X

Fonte: Dados do Estudo

Nota. Os espaços representados por um traço “-”, referem-se à ausência da variável que compõe o modelo dentro de cada classe.

A Tabela 3 apresenta o volume de casos que foram utilizados para a (BT) e (BP). Observa-se que, dentro do banco de dados a Microempresa é que contém o maior volume de casos e consequentemente

o maior volume de negociações da instituição. Outro fator de destaque está na quantidade de variáveis independentes selecionadas pelo método *stepwise*.

Tabela 3 - Número de casos válidos para amostra teste e principal

Percentagens	MEI	ME	PE	MédE	DG
20%	48	228	28	8	313
80%	212	852	87	28	1.178
100%	260	1080	115	36	1491
Percentagem em relação ao total de contratos	17,44%	72,43%	7,71%	2,41%	100,00%

Fonte: Elaborado pelos autores

4.6 Teste de multicolinearidade

Esse teste consiste no exame de correlação entre variáveis independentes, ocorrendo quando duas ou mais variáveis explicativas tentam explicar o mesmo fato (CUNHA; COELHO, 2017). A colinearidade pode ser mediada pela tolerância e sua inversa, chamada de Fator de Inflação de Variância (VIF), sendo

medidas bastante comuns para colinearidade (HAIR JR. *et al.*, 2009). A tolerância é calculada como $1 - R^{2*}$ e o VIF é calculado por meio do inverso da tolerância, quando o VIF for 1 e a tolerância for 1, implica dizer que não há multicolinearidade (HAIR JR. *et al.*, 2009). A Tabela 4 apresenta a estatística de colinearidade para as 4 classes de empresas e DG.

Tabela 4 - Teste dos Coeficientes de Colinearidade

Variáveis Independentes	MEI		ME		PE		MédE		DG	
	T*	VIF**	T*	VIF**	T*	VIF**	T*	VIF**	T*	VIF**
V2	-	-	-	-	-	-	-	-	0,978	1,022
V9	0,97	1,03	0,854	1,171	1	1	0,896	1,116	0,943	1,061
V19	0,97	1,03	0,914	1,095	-	-	0,9	1,111	0,981	1,019
V21	-	-	0,867	1,153	-	-	-	-	-	-
V27	-	-	0,972	1,029	-	-	-	-	-	-
V34	-	-	-	-	-	-	0,96	1,041	-	-
V35	-	-	0,881	1,135	-	-	-	-	-	-
V36	-	-	-	-	-	-	-	-	0,903	1,107

Fonte: Elaborado pelos autores, com utilização dos dados processados pelo software SPSS

Nota 1. *Tolerância,

Nota 2. **Fator de Inflação da Variância.

Analisando os resultados, os valores de tolerância ficaram muito próximos de 1, e o VIF também bastante próximo de 1 e distante de 10. Segundo Hair Jr. *et al.* (2009) uma referência de corte muito comum é um valor de tolerância de 0,10, o que corresponde a um valor VIF de 10.

4.7 Teste da Verossimilhança

O teste de *Log Likelihood Value*, objetiva estimar a probabilidade de um evento ocorrer, aferindo a capacidade do modelo (DIAS FILHO; CORRAR, 2017). O teste, também se mostra importante para verificar se o modelo apresenta melhora com a inclusão ou retirada de variáveis independentes, conforme apresentado na Tabela 5. Na RL, é estimado um modelo base, o qual tem a função de servir como

padrão para comparações, utilizando a soma dos quadrados das médias para estabelecer

o valor do logaritmo da verossimilhança $\{-2LL\}$ (HAIR JR. *et al.*, 2009).

Tabela 5 - Teste de Verossimilhança

MEI		MICRO		PEQUENA		MÉDIA	
E*	Verossimilhança de log -2	E*	Verossimilhança de log -2	E*	Verossimilhança de log -2	E*	Verossimilhança de log -2
1	169,284	1	640,259	1	57,707	1	177,214
2	149,379	5	585,234			3	150,313

Fonte: Elaborado pelos autores, com utilização dos dados processados pelo software SPSS

Nota. E* = Etapa

Valores menores de medida $-2LL$, melhoram o ajuste do modelo, sendo essa técnica utilizada pelo método *stepwise* para melhorar do passo (etapa) anterior, (HAIR JR. *et al.*, 2009). Observa-se que em todas as fases houve redução do log -2 para todas as etapas das classes de empresas. Outro instrumento utilizado para medição de modelos concorrentes foi *Nagelkerke*, os quais apresentaram os índices: 0,388, 0,491, 0,376 e 0,427 respectivamente a ordem de classificação das empresas.

A Tabela 6, apresenta os resultados para o teste de *Hosmer e Lemeshow*. Segundo Dias Filho e Corrar (2017), este

teste tem finalidade de verificar se existem diferenças significativas entre as classificações realizadas pelo modelo e as realidades observadas. Sua análise é com base na significância do modelo, o qual é favorável quando o nível de significância é igual ou superior a 0,05. Das classes apresentadas na Tabela 6, a única classe que não apresentou um nível significativo foi a Pequena Empresa, assim rejeitando a hipótese nula de não haver diferenças significativas. Entretanto, o método *stepwise* apresentou apenas uma etapa, não havendo outro nível para efetuar comparação.

Tabela 6 - Teste de *Hosmer e Lemeshow*

MEI				ME			
Etapa	Qui-quadrado	df	Sig.	Etapa	Qui-quadrado	df	Sig.
1	3,735	5	0,588	1	14,107	6	0,028
2	3,41	5	0,637	5	15,001	8	0,059
PE				MédE			
Etapa	Qui-quadrado	df	Sig.	Etapa	Qui-quadrado	df	Sig.
1	18,694	6	0,005	1	9,569	5	0,088
				3	10,731	7	0,151

Fonte: Elaborado pelos autores, com utilização dos dados processados pelo software SPSS

4.8 Modelos de pontuação de crédito (*credit scoring*) com a técnica estatística RL

As variáveis selecionadas para compor o modelo de *credit scoring* devem ter o poder de influenciar ou ter a possibilidade de influenciar um cliente a pender para o lado inadimplente ou adimplente na medida de sua influência. Conforme discorrem Dias Filho e Corrar (2017), “deve-se analisar o efeito que uma variável independente produz sobre a

dependente quando as demais se mantêm inalteradas”.

Quanto ao sinal das variáveis, ressalta-se que uma variação positiva em coeficiente negativo, sugere uma redução na probabilidade de inadimplência. Em caso de coeficiente ser positivo, sugere uma probabilidade de aumento da inadimplência.

A Tabela 7 apresenta os pesos atribuídos para cada variável independente incorporada ao modelo correspondente à classe de empresas selecionadas pelo método *stepwise*.

Tabela 7 - Equação Logística para MEI

Variáveis	Coefficiente	Equação Logística
V9	0,952	$\frac{1}{1 + e^{-(-1,274 + 0,952(V9) + 20,861(V19))}}$
V19	20,861	
(Constante)	-1,274	

Fonte: Elaborado pelos autores

O modelo, apresentado na Tabela 7, que compõe a Equação Logística para as empresas classificadas como MEI, destaca a

inclusão de 2 variáveis, sendo uma primitiva (V9) e outra artificial (V19).

Tabela 8 - Equação Logística para ME

Variáveis	Coefficiente	Equação Logística
V9	0,942	$\frac{1}{1 + e^{-(-1,734 + 0,942(V9) + 2,222(V19) + 0,628(V21) - 1,020(V27) + 0,744(V35))}}$
V19	2,222	
V21	0,628	
V27	-1,02	
V35	0,744	
(Constante)	-1,734	

Fonte: Elaborado pelos autores

O modelo, apresentado na Tabela 8, que compõe a Equação Logística para as empresas classificadas como ME, destaca a

inclusão de 5 variáveis, sendo uma primitiva (V9) e 4 variáveis artificiais (V19), (V21), (V27) e (V35)

Tabela 9 - Equação Logística para PE

Variáveis	Coefficiente	Equação Logística
V9	0,753	1
(Constante)	-0,608	$1 + e^{-(-0,608 + 0,753 (V9))}$

Fonte: Elaborado pelos autores

O modelo, apresentado na Tabela 9, que compõe a Equação Logística para as empresas classificadas como PE, destaca a inclusão de apenas uma variável, sendo ela primitiva (V9). A seleção de apenas uma variável para compor o modelo, está de

acordo com a questão acadêmica devido a respeitar as premissas da técnica, porém pode dificultar sua aceitação em nível de gestão. O aprimoramento do modelo com a inclusão de variáveis de gestão ou de restrição de risco podem melhorar a confiabilidade no modelo.

Tabela 10 - Equação Logística para MédE

Variáveis	Coefficiente	Equação Logística
V9	0,753	1
V19	21,013	
V34	2,284	
(Constante)	-1,505	$1 + e^{-(-1,505 + 0,753 (V9) + 21,013 (V19) + 2,284 (V34))}$

Fonte: Elaborado pelos autores

O modelo, apresentado na Tabela 10, que compõe a Equação Logística para as empresas classificadas como MédE, destaca

a inclusão de 3 variáveis, sendo uma primitiva (V9) e 2 variáveis artificiais (V19) e (V34).

Tabela 11 - Equação Logística para DG

Variáveis	Coefficiente	Equação Logística
V2	-0,229	1
V9	0,909	
V19	2,057	
V26	-1,106	
V36	-1,099	
(Constante)	-0,031	$1 + e^{-(-0,031 - 0,229 (V2) + 0,909 (V9) + 2,057 (V19) - 1,106 (V26) - 1,099 (V36))}$

Fonte: Elaborado pelos autores

O modelo, apresentado na Tabela 11, que compõe a Equação Logística para as empresas classificadas como DG, destaca a inclusão de 5 variáveis, sendo duas primitivas (V2), (V9) e 3 variáveis artificiais (V19), (V26) e (V36).

Pode-se observar que o modelo de DG, por meio do método *stepwise*, incluiu o máximo de 5 variáveis, apresentando o

mesmo número de variáveis do modelo da ME, sendo esta, a classe que apresenta maior volume de dados.

4.9 Discussão do poder discriminatório da RL

Nos Quadros de 8 a 12, são apresentados os resultados dos testes de

acurácia da BT, e da BP para cada uma das classes de empresas e DG. No Quadro 13 é ilustrado o resumo da acurácia dos testes, já discutido nos Quadros 8 a 12, juntamente com o teste de acurácia sobre os DG, envolvendo 1.491 casos.

Ao observar o Quadro 8 os resultados do teste de acurácia para MEI, apresentaram uma superioridade de 3,8% pontos percentuais, da BP sobre a BT, mas

ambos os indicadores acima de 65%. A acurácia geral atingiu percentual de 83,0%, no qual é interessante observar que a predição para o adimplente atingiu percentagem de 93,6%, apresentando maior poder discriminatório para identificação do adimplente. Devido à baixa quantidade de dados que compôs a amostra teste, não foram selecionadas observações para inadimplentes na BT.

Quadro 8 - Resultados dos testes de acurácia para classe de empresas classificadas como MEI

Acurácia	MEI Base Principal			MEI Base Teste		
	Inadimplente	Adimplente	%	Inadimplente	Adimplente	%
Inadimplente	16	25	39,0	0	10	0
Adimplente	11	160	93,6	0	38	100
Percentagem Global			83,0			79,2

Fonte: Dado do estudo

No Quadro 9, a acurácia para a classe de ME entre a BT e BP apresenta uma pequena diferença de 2,8% pontos percentuais, esta proximidade demonstra uma boa adaptabilidade do modelo, com

acurácia geral de 84,9%. A Microempresa, na BP, apresentou um poder de previsão, para o inadimplente de 64,1%, e uma acurácia bastante significativa para o adimplente de 91,1%.

Quadro 9 - Resultados dos testes de acurácia para as empresas classificadas como ME

Acurácia	ME Base Principal			ME Base Teste		
	Inadimplente	Adimplente	%	Inadimplente	Adimplente	%
Inadimplente	125	70	64,1	34	14	70,8
Adimplente	59	603	91,1	14	166	92,2
Percentagem Global			84,9			87,7

Fonte: Dados do estudo

O Quadro 10 ilustra os resultados da acurácia para as empresas classificadas na classe de PE. O nível de acurácia entre a BT e BP apresentou uma diferença de 5,2% pontos percentuais, demonstrando a capacidade de predição do modelo, sobre a BP de 88,5%, composta de 212 contratos. A

percentagem de acerto, sobre o inadimplente, deve ser analisada com cuidado, devido a apresentar percentual abaixo de 50%, este fato pode estar relacionado com o volume de contratos incluídos no teste.

Quadro 10 - Resultados dos testes de acurácia para classe de empresas classificadas como PE

Acurácia	PE Base Principal			PE Base Teste		
	Inadimplente	Adimplente	%	Inadimplente	Adimplente	%
Inadimplente	6	9	40	4	6	40,0
Adimplente	1	71	98,6	2	36	94,7
Percentagem Global			88,5			83,3

Fonte: Dados do estudo

O Quadro 11 ilustra a acurácia obtida para a classe da MédE. O volume de contratos utilizados na BP foi de 227 contratos, apresentando uma acurácia de 83,3%. A BT apresentou uma acurácia de

100% tanto para adimplentes como para inadimplentes, entretanto, o volume de contratos que compôs a BT foi de apenas 12, o que pode distorcer os valores observados.

Quadro 11 - Resultados dos testes de acurácia para as empresas classificadas como MédE

Acurácia	MédE Base Principal			MédE Base Teste		
	Inadimplente	Adimplente	%	Inadimplente	Adimplente	%
Inadimplente	15	28	35	2	0	100
Adimplente	10	174	94,6	0	10	100
Percentagem Global			83,3			100

Fonte: Dados do estudo

Posteriormente à aplicação da técnica da RL, nas classes: MEI, ME, PE e MédE, sobre o volume de contratos/clientes, utilizou-se a base geral, aqui denominada de DG, com 1.491 observações válidas, nos quais foram aplicados os mesmos procedimentos das classes de empresas.

Os resultados estão ilustrados no Quadro 12, sendo que a BT e a BP

apresentaram acurácias praticamente iguais, ou seja 85,4% e 85,0%, respectivamente. Constatou-se boa capacidade de previsão do modelo, mostrando, principalmente que, a acurácia destinada a prever o adimplente, teve bom desempenho. Ao analisar o inadimplente, o desempenho foi apenas regular, alcançando 59% e 52,5%.

Quadro 12 - Resultados dos testes de acurácia para a base de DG.

Acurácia	DG Base Principal			DG Base Teste		
	Inadimplente	Adimplente	%	Inadimplente	Adimplente	%
Inadimplente	134	121	52,5	49	34	59,0
Adimplente	55	861	94,0	12	220	94,8
Percentagem Global			85,0			85,4

Fonte: Dados do estudo

Na observação do Quadro 13, os DG ficaram com percentuais próximos ao das classes de empresas, principalmente MEI e ME, porém, mesmo pequenas variações

percentuais podem representar valores significantes. Uma observação importante é que os modelos com acurácia similares são aqueles que possuem as maiores bases de

dados, entretanto as classes da PE e MédE são as classes que possuem menores distâncias dentro da base de dados,

conforme pode-se observar nas discretizações.

Quadro 13 - Comparativo da acurácia das classes de empresas com a base de DG

CLASSES	MEI	ME	PE	MédE	DG
Acurácia	83%	84,9%	88,5%	83,0%	85,0%

Fonte: Dados do estudo

Pode-se observar que a RL apresenta uma boa aderência aos pressupostos. Para Dias Filho e Corrar (2017), o modelo logístico acolhe com mais facilidade variáveis categóricas, sendo uma das razões que se torna boa alternativa à Análise Discriminante.

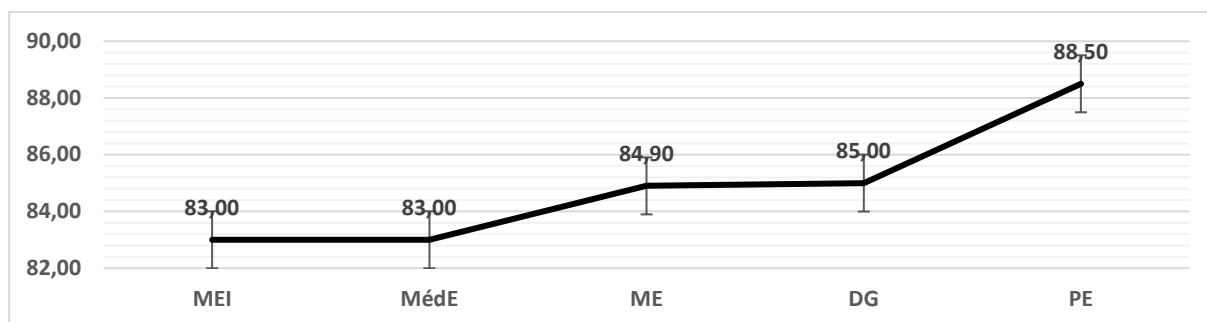
Ao analisar o Quadro 13 verifica-se que a classe de empresas PE, apresenta a melhor percentagem preditiva, porém ao observar a composição das variáveis do modelo para PE, conforme Tabela 2, apenas uma variável foi selecionada pelo método *stepwise*. Para uso prático do modelo para PE, recomenda-se que outras variáveis, não contempladas neste estudo, sejam testadas para inclusão, objetivando dar maior robustez ao modelo.

Na Tabela 3, também é possível observar que a classe de empresas ME é a que mais possui vínculo de negócios na

carteira de crédito com a Instituição Financeira, representando 72,43% do total dos contratos. Esse percentual demonstra que a grande maioria dos clientes da instituição, apresentam-se com faturamento dentro das características da ME. Devido a esse fator, a acurácia dessa técnica torna-se importante para análise da carteira de clientes da instituição, a qual performou em 84,9%.

Observou-se que as percentagens da acurácia das 4 classes de empresas: MEI, ME, PE e MédE, são próximas, porém, não iguais à dos DG. O Quadro 14 ilustra a distribuição da acurácia em ordem crescente, onde pode-se observar que as classes MEI, MédE e ME encontram-se até 2 pontos percentuais abaixo dos DG e a classe PE encontra-se 3,5 pontos percentuais acima dos DG.

Quadro 14 - Distribuição da Acurácia, em ordem crescente.



Fonte: Elaborado pelos autores

ALVIM, P. C. R. C. O papel da informação no processo de capacitação tecnológica das micro e pequenas

empresas. **Ciência da Informação**, Brasília, v.27, n.1, p. 28-35, 1998.

ARAÚJO, E. A.; CARMONA, C. U. M. Construção de modelos *credit scoring* com Análise Discriminante e Regressão Logística para a gestão do risco de inadimplência de uma instituição de microcrédito. **Revista Eletrônica de Administração**, v.15, n.1, 2009.

BANCO NACIONAL DE DESENVOLVIMENTO - BNDES. *Classificação de porte de empresa*. Disponível em: <https://www.bndes.gov.br/wps/portal/site/home/imprensa/noticias/conteudo/20100622_modificacao_porte_empresa>. Acesso em: 11 nov 2017.

BERGER, N. A.; COWAN, A. M.; FRAME, W. S. The surprising use of credit scoring in small business lending by community banks and the attendant effects on credit availability, risk, and profitability. **Journal of Financial Services Research**, v.39, n.1, p. 1–17, 2011.

BERTI, A. Consultoria e Diagnóstico Empresarial. 2.ed. Curitiba: Juruá, 2012.

BRASIL, Lei nº 128, de 19 de dezembro de 2008, **Presidência da República. Casa Civil. Subchefia para Assuntos Jurídicos teoria e prática**. Disponível: <http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp128.htm>. Acesso em: 12 ago 2017.

BRASIL, Lei nº 123, de 14 de dezembro de 2006, **Presidência da República. Casa Civil. Subchefia para Assuntos Jurídicos teoria e prática**. Disponível: http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp123.htm. Acesso em: 12 ago 2017.

BRUNI, A. L. **A análise contábil e financeira**. 4. ed. São Paulo: Atlas. 2010.

CAMARGOS, M. A. DE.; CAMARGOS, M. C. S.; ARAÚJO, E. A. A inadimplência em um programa de crédito de uma instituição financeira pública de minas gerais: uma análise utilizando Regressão Logística, **REGE**, São Paulo, v.19, n.3, p. 467–486, 2012.

CIAMPI, F.; GORDINI, N. Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. **Journal of Small Business Management**, v.51 n.1, p.23-45, 2013.

CUNHA, J. V. A.; COELHO, A. C. Regressão Linear Múltipla. *In*: Corrar, L. J. (Coord.); Paulo, E. (Coord.); Dias Filho, J. M. (Coord.). **Análise Multivariada: para os Cursos de Administração, Ciências Contábeis e Economia**. São Paulo: Atlas, cap. 3, 2017.

DIAS FILHO, J. M.; CORRAR, L. J. Regressão Logística. *In*: Corrar, L. J. (Coord.); Paulo, E. (Coord.); Dias Filho, J. M. (Coord.). **Análise Multivariada: Para os Cursos de Administração, Ciências Contábeis e Economia**. São Paulo: Atlas, cap. 5, 2017.

DINIZ, C.; LOUZADA, F. Métodos Estatísticos para Análise de Dados de Crédito. *In*: 6th BRAZILIAN CONFERENCE ON STATISTICAL MODELLING IN INSURANCE AND FINANCE. Maresias – São Paulo, ABE, USP, UNICAMP 24 a 28 de março 2013.

DUARTE, S. V.; FURTADO, M. S. V. **Trabalho de conclusão de curso em Ciências Sociais Aplicadas**. 1.ed. São Paulo: Saraiva, 2014.

GARCÍA S. *et al.* A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. **IEEE Transactions on Knowledge and Data Engineering**, v.25 n.4, p.734-750, 2013.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social**, 7.ed. São Paulo: Atlas, 2009.

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**. UFSC, Florianópolis, v.10, n.20, p.139-160, 2013

GOUVÊA, M. A.; GONÇALVES, E. B. Análise de risco de crédito com o uso de modelos de Regressão Logística e redes neurais. **Globalização e Internacionalização de Empresas**, X SEMINÁRIO EM ADMINISTRAÇÃO FEA-USP, 09 a 19 de agosto 2007.

HAIR JR, *et al.* **Análise multivariada de dado**. Tradução Adonai Schlup Sant'Anna. 6.ed. Porto Alegre: Bookman, 2009.

KELLY, M. G.; HAND, D. J. Credit scoring with uncertain class definitions. **IMA Journal of Management Mathematics**, v. 10, n. 4, p. 331-345, 1999.

LEMOS, E. P.; STEINER, M. T. A.; NIEVOLA, J. C. Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. **Revista de Administração**. São Paulo, v.40, n.3, p.225-234, jun./ago./set, 2005.

LI, K.; NISKANEN, J.; KOLEHMAINEN, M. Financial innovation credit default hybrid model for SME lending. **Expert Systems With Applications**, v.61 p. 343-355, 2016.

LOUZADA, F.; ARA, A. FERNANDES, G. B. Classification methods applied to credit scoring: Systematic review and overall comparison. **Surveys in Operations Research and Management Science**. v.21 n.2, p.117– 134, 2016.

LUNET, N.; SEVERO, M.; BARROS, H. Desvio padrão ou erro padrão: Notas Metodológicas. ArquiMed, 2006.

Disponível:

<http://www.scielo.mec.pt/pdf/am/v20n1-2/v20n1-2a08.pdf>_ Acesso em: 15 jan. 2018.

LUO, C.; WU, Desheng.; WU, Dexiang. A deep learning approach for credit scoring using credit default swaps. **Engineering Applications of Artificial Intelligence**, v.65, p 465-470, October 2017.

LUO, S.; KONG, X.; NIE, T. Spline based survival model for credit risk modeling. **European Journal of Operational Research**, v.253. n.3, p 869-879, September 2016.

MARQUÉS, A. I.; GARCÍA, V.; SÁNCHEZ, J. S. Exploring the behaviour of base classifiers in credit scoring ensembles. **Expert Systems with Applications**, v.39 n.11, p.10244–10250, September 2012.

MARQUES, J. M.; LIMA, J. D. de. A estatística multivariada na análise econômico-financeira de empresas. **Revista FAE**, Curitiba, v.5, n.3, p.51-59, 2002.

MILERIS, R. Macroeconomic Determinants of Loan Portfolio Credit Risk in Banks. **Inzinerine Ekonomika-Engineering Economics**, v.23 n.5, p.496–504, 2012.

MISSIO, F.; JACOBI, L. Variáveis dummy: especificações de modelos com parâmetros variáveis. **Ciência e Natura**, v.29, n.1, p.111-135. 2007.

MSELMY, N.; LAHIANI, A.; HAMZA, T. Financial distress prediction: The case of French small and medium-sized firms. **International Review of Financial Analysis**. v.50 p. 67-80, 2017.

OHLSON, J. A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. **Journal of Accounting Research**, v.18 n.1, p.109-131, 1980.

PRADO *et al.* Multivariate analysis of credit risk and bankruptcy research data: a bibliometric study involving different knowledge fields (1968–2014). UFLA. **Scientometrics**. v.106, n.3, p. 1007–1029, 2016.

RODRIGUES, A.; PAULO, E. Introdução à Análise Multivariada. *In*: Corrar, L. J. (Coord.); Paulo, E. (Coord.); Dias Filho, J. M. (Coord.). **Análise Multivariada**: para Curso de Administração, Ciências Contábeis e Economia. São Paulo: Atlas, cap. 1, 2017.

ROGERS, Dany.; MENDES-DASILVA, W.; Rogers Pablo. Credit Rating Change and Capital Structure in Latin America. **Brazilian Administration Review - BAR**, v.13, n2, p.1–22, 2016.

SELAU, L. P. R.; RIBEIRO, J. L. D. Uma sistemática para construção e escolha de modelos de previsão de risco de crédito. **Gestão e Produção**, v.16, n.3, p.398-413, 2009.

SELAU, L. P. R.; RIBEIRO, J. L. D. A systematic approach to construct credit risk

forecast models. **Pesquisa Operacional**, v.31 n.1, p. 41–56, 2011.

SILVA, R. A.; RIBEIRO, E. M.; MATIAS, A. B. Aprendizagem estatística aplicada à previsão de default de crédito. **Revista de Finanças Aplicadas**, v.7, n.2, p.1-19, 2016.

SILVA, J. P. D. **Análise Financeira das Empresas**. São Paulo: Atlas, 2004.

SILVA, J. P. D. **Gestão e Análise de Risco de Crédito**. 9.ed. rev. e atualizada. São Paulo: Cangage Learning, 2016.

SMARANDA, C. Scoring Functions and Bankruptcy Prediction Models: Case Study for Romanian Companies. **Procedia Economics and Finance**, v.10, p.217-226, 2014.

VENTURA, R. Mudanças no Perfil do Consumo no Brasil; Principais Tendências nos Próximos 20 Anos. **Macroplan – Prospectiva, Estratégia e Gestão**, agosto 2010.

ZHU, Y. *et al.* Predicting China's SME credit risk in supply chain financing by logistic regression, artificial neural network and hybrid models. **Sustainability (Switzerland)**, v.8, n.5, p.1-17, 2016.

ⁱ Mestrado concluído em 2018 no Programa de Pós-Graduação de Engenharia de Produção. Professor do Curso de Ciências Contábeis no IFPR-Campus Palmas. Departamento/Colegiado de Ciências Contábeis. Área de Sociais Aplicadas.

ⁱⁱ Doutorado em Engenharia de Produção pela Universidade Federal do Rio Grande do Sul (UFRGS) Professor do programa de Pós-graduação em Engenharia de Produção e Pesquisador da UTFPR.

ⁱⁱⁱ Doutor em Administração pela Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (2009). Professor dos programas de pós-graduação em Desenvolvimento Regional e Engenharia de Produção e Sistemas.

^{iv} Doutor em Ciências Contábeis e Administração (2015). Atualmente é professor titular C, atuando no Mestrado em Ciências Contábeis e Administração da Unochapecó, na graduação e na especialização.