

Classificação de risco de crédito: um comparativo entre modelos de análise discriminante e regressão logística nas empresas de capital aberto brasileiras

Fabrício de Macedo
Centro Universitário Católica de SC
(fabricio.macedo@catolicasc.org.br.)

Stefani da Silva
Centro Universitário Católica de SC
(stefani.silva@catolicasc.org.br.)

Recebido em 06/02/2019
Aprovado em 16/02/2020

Resumo

Os modelos de classificação de risco de crédito podem ser utilizados pelas empresas para prever uma situação financeira difícil com antecedência, evitando ou reduzindo os prejuízos financeiros. Também, contribuem para a reversão da insolvência, uma vez que um modelo de previsão eficaz pode prever problemas financeiros com antecedência, desta forma, a empresa possui tempo para tomada de decisões e ações necessárias. Sendo assim, o objetivo deste estudo é identificar o modelo mais adequado para classificação de risco de crédito das companhias abertas brasileiras, “comparando modelos” de análise discriminante e regressão logística de vários autores já realizaram. Foram utilizados como procedimentos metodológicos os métodos de pesquisa bibliográfica e documental com uma abordagem quantitativa. A amostra da pesquisa compreende 38 companhias abertas brasileiras. Os modelos escolhidos para validação foram os de Altman, Baidya e Dias; Matias; Brito e Assaf Neto e Scarpel e Milioni. Utilizou-se o *software SPSS Statistics* para realizar os testes de qualidade e significância dos modelos e a construção da curva ROC para validação. Os resultados do estudo indicam que a técnica de regressão logística apresenta um ótimo poder de discriminação, demonstrando uma área de 0,911 na curva ROC, podendo prever a insolvência com antecedência com um nível de 78,9% de precisão.

Palavras-chave: Risco de crédito; Insolvência; Análise discriminante; Regressão logística

Abstract

Credit risk classification models can be used by companies to predict a difficult financial situation in advance, avoiding or reducing financial losses. Also, they contribute to the reversal of insolvency, since an effective forecasting model can foresee financial problems in advance, in this way, the company has the time to make necessary decisions and actions. Therefore, the objective of this study is to identify the most suitable model for credit risk classification of Brazilian public companies, "comparing models" of discriminant analysis and logistic regression of several authors already performed. Methods of bibliographic and documentary

research with a quantitative approach were used as methodological procedures. The research sample comprises 38 Brazilian public companies. The models chosen for validation were those of Altman, Baidya and Dias; Matias; Brito and Assaf Neto and Scarpel and Milioni. The SPSS Statistics software was used to carry out the tests of quality and significance of the models and the construction of the ROC curve for validation. The results of the study indicate that the logistic regression technique presents an excellent discriminating power, showing an area of 0.911 in the ROC curve, and can predict insolvency in advance with a level of 78.9% accuracy. **Keywords:** Credit risk, Insolvency, Discriminant analysis, Logistic regression.

1. INTRODUÇÃO

O processo de gerenciamento de risco de crédito pode ser definido, de acordo com Caouette, Altman e Narayanan (1998), como a possibilidade de que a expectativa de recebimento de uma quantia, durante um determinado período, não se cumpra. Conforme Bessis (2011), o risco de crédito pode ser entendido como as perdas geradas por um evento de *default* do tomador ou pela redução da sua qualidade de crédito. Sendo assim, há diversas situações que caracterizam um evento de *default*, como por exemplo, o atraso no pagamento de uma obrigação ou o início de um procedimento legal como a falência.

Diante disso, Minussi, Damacena e Ness Junior (2002) explicam que, ao aplicar um modelo de previsão de insolvência, estima-se que os problemas de uma companhia podem ser previamente identificados, uma vez que sua situação financeira tende a agravar, até atingir o estágio de falência.

Martins (2003) afirma que os primeiros estudos relacionados à previsão de insolvência foram desenvolvidos por volta da década de 30, no entanto, somente a partir da década de 60, com o uso de técnicas estatísticas é que o tema ganhou impulso. Minussi, Damacena e Ness Junior (2002) citam que existem diversas técnicas multivariadas para a construção de modelos de previsão de insolvência. Dentre elas, a análise discriminante e a regressão logística são as mais utilizadas (MINUSSI; DAMACENA; NESS Junior, 2002).

Diante deste contexto, este estudo procura responder a seguinte questão de pesquisa: Qual o modelo mais adequado para classificação de risco de crédito das companhias abertas brasileiras utilizando um comparativo entre os modelos de análise discriminante e regressão logística de vários autores já realizaram?

Para tanto, possui como objetivos específicos realizar o levantamento e diagnosticar a produção científica sobre risco de crédito, técnicas estatísticas e modelos de classificação de crédito, avaliar os modelos de classificação de risco de crédito a partir dos indicadores contábeis, aplicar as variáveis na amostra selecionada por meio dos testes de qualidade e significância e mensurar o poder discriminante dos modelos através da curva ROC, identificando o mais adequado.

De acordo com Stüpp (2015), a previsão de insolvência pode ser utilizada pelos usuários das empresas como uma ferramenta para analisar o risco de crédito e prever uma situação financeira difícil com antecedência, podendo evitar ou reduzir os prejuízos financeiros.

Um estudo realizado pela Serasa Experian identificou que no ano de 2016 houve um recorde histórico em recuperações judiciais requeridas com 1.770 solicitações, demonstrando um aumento de 44,8% em relação a 2015. O número de pedidos de falência também é preocupante, visto que neste ano cerca de 1.846 empresas entraram com o processo. (SERASA EXPERIAN, 2017).

O setor de serviços registrou o maior número de pedidos de recuperação judicial do ano com 713 solicitações, demonstrando um aumento de 48,5% em relação a 2015. Em seguida, o setor de comércio apresentou 611 pedidos, tendo um crescimento de 51,2% em relação ao ano anterior, enquanto a indústria obteve um aumento de 24,2% com 446 pedidos de recuperação judicial (SERASA EXPERIAN, 2017).

Esta pesquisa justifica-se pela contribuição aos usuários das entidades, visto que visa garantir o modelo mais adequado para classificação de risco de crédito e, assim, auxiliar os gestores e investidores das empresas na tomada de decisões com uma ferramenta sólida. Também contribui para a reversão da insolvência, uma vez que, um modelo de previsão eficaz pode prever uma situação financeira difícil com antecedência.

Para tanto, a estrutura deste artigo está organizada, além desta introdução, em quatro seções: na próxima seção, é apresentada a fundamentação teórica, que discorre sobre os principais pontos do risco de crédito, as técnicas estatísticas, os modelos de classificação de risco de crédito e estudos correlacionados. Já a terceira discorre sobre a metodologia utilizada no estudo: a pesquisa documental se deu por meio da análise das demonstrações contábeis das empresas, os testes de qualidade e significância foram utilizados para análise das funções e a curva ROC para validar os modelos analisados. Em seguida, a quarta seção apresenta a análise e validação dos modelos de previsão de insolvência. Por fim, finaliza-se com a conclusão da pesquisa e suas recomendações para estudos futuros.

2. Fundamentação Teórica

Neste tópico são apresentados os temas para o desenvolvimento deste estudo, tais como o risco de crédito, técnicas estatísticas, modelos de classificação de risco de crédito e estudos correlacionados, seus conceitos e suas aplicações.

2.1 RISCO DE CRÉDITO

Segundo Brito e Assaf Neto (2008), o risco de crédito pode ser compreendido como a probabilidade de o credor ter prejuízos financeiros em razão de o tomador não quitar as obrigações assumidas nas condições acordadas. Sendo assim, Minussi, Damacena e Ness Junior (2002) afirmam que a análise de risco de crédito requer a definição minuciosa de critérios que possam indicar a possibilidade de inadimplência do tomador.

De acordo com Davis (1999), existem três modelos de análise de risco de crédito que buscam auxiliar os gestores na tomada de decisão de crédito, são eles: modelos preditivos, de risco e de *default*.

Os modelos preditivos são aplicados para previsão da capacidade de um cliente quitar suas obrigações, geralmente é utilizado para aprovação de um crédito ou para aumento de limite. Por outro lado, os modelos de risco são utilizados para prever se clientes que já fazem parte da carteira da empresa tendem a pagar ou estender uma possível inadimplência no pagamento, o principal objetivo deste modelo é analisar seu potencial de coleta. Por fim, os modelos de *default* são utilizados para prever se o cliente tem tendência à falência (DAVIS, 1999).

2.2 TÉCNICAS ESTATÍSTICAS

Conforme Mário (2002), busca-se verificar com a utilização de técnicas estatísticas se determinada empresa se classifica no grupo de solventes ou insolventes por meio da pontuação adquirida com base em dados obtidos das demonstrações contábeis. Assim, através do estudo realizado por Stüpp (2015), não há uma definição quanto à técnica estatística mais

aplicada pelos autores para previsão de insolvência, contudo, dentre as mais utilizadas, destaca-se a análise discriminante e a regressão logística.

Nesse sentido, Brito e Assaf Neto (2008) afirmam que a regressão logística apresenta vantagens em relação à análise discriminante, uma vez que suas suposições iniciais são mais flexíveis. Pode-se observar que a análise discriminante está baseada em requisitos mais limitados, como a normalidade das variáveis independentes e igualdade das matrizes de variância-covariância dos grupos.

2.2.1 Análise Discriminante

A análise discriminante é uma técnica estatística que procura definir dois ou mais grupos e classificar as observações em um deles com base em suas características individuais. É uma técnica utilizada para classificações nas quais a variável dependente é categórica, ou seja, qualitativa, como por exemplo: masculino ou feminino, solvente ou insolvente e as variáveis independentes são quantitativas (SANVINCENTE; MINARDI, 1998). Nessa perspectiva, Hair et al. (2009) ressalta que a análise discriminante busca realizar a combinação linear de duas ou mais variáveis independentes que discriminará da melhor forma as observações nos grupos definidos inicialmente.

De acordo com Stüpp (2015), a variável estatística para a análise discriminante, também é denominada função discriminante, que é definida a partir de uma equação para ampliar a discriminação dos grupos. A equação é representada da seguinte forma: $Z = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$. Em que: Z é a variável dependente categórica, representada por uma pontuação; a é o intercepto da função; b é o coeficiente discriminante para a variável independente e X que é um vetor de covariáveis “tem por finalidade permitir que os animais sejam comparados e avaliados em igualdade de condições para estes efeitos, pois influem diretamente sobre o desempenho no período de teste dos indivíduos” (PITA; ALBUQUERQUE, 2001, p, 2).

Barth (2002) explica que, após definida a função discriminante, deve-se definir o ponto de corte para utilizá-lo na classificação das observações iniciais para duas classes de análise. O ponto de corte é definido partindo do princípio de que, se Z for maior ou igual ao ponto de corte, a observação será classificada no grupo 1 (Primeira classe), por outro lado, se Z for menor que o ponto de corte, a observação será classificada no grupo 2 (Segunda Classe).

2.2.2 Regressão Logística

A regressão logística, também conhecida como análise *logit*, é definida por Hair et al. (2009) como uma forma especializada de regressão formulada para prever e explicar uma variável categórica binária, e variáveis independentes numéricas e não numéricas.

De acordo com Barth (2002), o modelo de previsão utilizado na regressão logística é definido como: $\ln\left(\frac{p}{1-p}\right) = b_0 + b_1.X_1 + b_2.X_2 + b_3.X_3 + \dots + b_n.X_n$. Em que: p é a probabilidade de solvência; $(1 - p)$ é a probabilidade de insolvência; X sendo um vetor de covariáveis são as variáveis preditoras; b sendo um vetor de betas são os coeficientes a serem atribuídos a cada uma das variáveis X sendo um vetor de covariáveis.

Conforme Stüpp (2015), na regressão logística, os coeficientes estimam o efeito das alterações nas variáveis independentes sobre o logaritmo natural da razão de probabilidades, conhecido com *logit*. Para avaliar o impacto dos fatores sobre a probabilidade de solvência, eles devem ser transformados por meio de antilogaritmo, que é obtido pela expressão:

$$p = \frac{1}{1 + e^{b_0 + b_1.X_1 + b_2.X_2 + b_3.X_3 + \dots + b_n.X_n}}$$

Conforme Brito e Assaf Neto (2008), através dos resultados da regressão logística é possível medir a probabilidade de determinada empresa apresentar um estado de insolvência ou solvência, com base em um conjunto de características.

2.3 MODELOS DE CLASSIFICAÇÃO DE RISCO DE CRÉDITO

Segundo Andrade (2004), existem três categorias de modelos de risco de crédito: modelos estocásticos de risco de crédito, modelos de risco de portfólio e modelos de classificação de risco. Os modelos estocásticos buscam avaliar o comportamento estocástico do *default* e as variáveis que o determinam, os modelos de risco de portfólio têm por objetivo quantificar os riscos de uma carteira de crédito, medindo a distribuição de perdas ou de valor dela.

Por outro lado, os modelos de classificação de risco visam avaliar o risco de um devedor ou operação específica, atribuindo uma medida que corresponde à expectativa de ocorrência de *default*. Esse indicador pode ser expresso na forma de *rating* ou *scoring* (ANDRADE, 2004).

Chaia (2003) afirma que o objetivo do *scoring* é identificar os aspectos que definem a probabilidade de inadimplência de determinada empresa por meio da utilização de técnicas estatísticas. Em contrapartida, Mendes (2014) afirma que a avaliação de *rating* contempla aspectos qualitativos, incluindo os fatores específicos da entidade e as características do setor e os aspectos quantitativos que têm foco na análise financeira.

2.3.1 Modelo de Altman, Baidya e Dias

Altman, Baidya e Dias (1979) desenvolveram um modelo de previsão de insolvência utilizando a análise discriminante com uma amostra composta por 58 empresas, sendo 23 com problemas financeiros e 35 empresas de setores semelhantes e de porte equivalente às empresas da primeira amostra. O modelo obteve cerca de 84,2% de acerto na classificação das empresas e possui a seguinte equação: $Z = 1,44 + 4,03x_2 + 2,25x_3 + 0,14x_4 + 0,42x_5$.

As variáveis são obtidas pelas seguintes fórmulas: x_2 : (não exigível - capital aportado pelos acionistas) / ativo total, x_3 : lucro antes dos juros e impostos / ativo total, x_4 : patrimônio líquido / exigível total e x_5 : vendas / ativo total. O ponto de corte definido desse modelo é 0. (ALTMAN et al., 1979).

2.3.2 Modelo de Matias

De acordo com Birolo, Cittadin e Ritta (2011) e Matias (1978), o modelo de Matias foi desenvolvido em 1978 por meio da análise discriminante a partir de uma mostra de 50 empresas. O modelo é apresentado pela seguinte equação: $Z = 23,792x_1 - 8,260x_2 - 8,868x_3 - 0,764x_4 + 0,535x_5 + 9,912x_6$. Em que: x_1 : patrimônio líquido/ativo total, x_2 : financiamentos e empréstimos bancários/ativo circulante, x_3 : fornecedores/ativo total, x_4 : ativo circulante/passivo circulante, x_5 : lucro operacional/lucro bruto e x_6 : disponível/ativo total.

2.3.3 Modelo de Brito e Assaf Neto

O modelo de previsão de insolvência de Brito e Assaf Neto (2008) foi desenvolvido com a utilização da regressão logística, tendo como amostra 30 empresas solventes e 30 insolventes. É apresentado pela seguinte equação: $\ln\left(\frac{p}{1-p}\right) = -4,740 - 4,528x_{12} +$

$18,433_{x16} - 414,080_{x19} - 11,028_{x22}$. Em que: $x12$: (lucros acumulados - reserva de lucros) / ativo total, $x16$: (passivo circulante financeiro + exigível a longo prazo financeiro) / ativo total, $x19$: (ativo circulante - passivo circulante) / ativo total e $x22$: (ativo circulante financeiro - passivo circulante financeiro) / vendas líquidas. O modelo obteve 88,3% de classificações corretas e possui ponto de corte 0.

2.3.4 Modelo de Scarpel e Milioni

De acordo com Scarpel e Milioni (2002) seu modelo foi desenvolvido através da regressão logística a partir de uma amostra de 60 empresas brasileiras de capital aberto, a eficiência global foi de 87%. O modelo é apresentado pela seguinte equação: $ln\left(\frac{p}{1-p}\right) = 0,190 + 3,312_{GAi} - 3,687_{EGi} + 10,997_{ROAAJIR}$. Em que: giro do ativo total (GA): receita/ativo total, endividamento geral (EG): exigível total/patrimônio líquido, retorno sobre o ativo total antes do pagamento de juros e imposto (ROA AJIR): lucro antes do pagamento de juros e imposto de renda/ativo total (SCARPEL; MILIONI, 2002).

2.5 ESTUDOS CORRELACIONADOS

Para Mendes (2014) afirma que o primeiro estudo de insolvência nas empresas é o de Fitzpatrick, desenvolvido em 1932. Esta pesquisa analisou 19 empresas solventes e insolventes, entre 1920 e 1929, comparando os índices dos dois grupos.

Pinheiro et al. (2007) cita que os estudos realizados sobre previsão de insolvência, tanto no Brasil como no exterior, fazem o uso de diferentes métodos estatísticos, sendo as principais técnicas utilizadas a Análise Discriminante Linear, Regressão Logística e Redes Neurais.

Silva (1982) apresenta modelos para classificação de empresas, quanto ao estado de solvência, para fins de concessão de crédito, obtendo como resultado um modelo desenvolvido em 1982, baseado na Análise Discriminante. Para operacionalizar o estudo a amostra foi composta por 419 empresas, divididas em empresas solventes e insolventes, sendo 337 empresas industriais e 82 comerciais. O modelo de Silva para a indústria obteve 90% de acerto na classificação de empresas solventes e 83% na classificação das insolventes. Já o modelo desenvolvido para empresas comerciais alcançou 90% de acerto na classificação das solventes e 72% na classificação das insolventes.

Em seguida, Sanvicente e Minardi (1998) utilizaram uma amostra constituída por 92 empresas com ações negociadas na Bolsa de Valores de São Paulo, dentre as quais 46 tiveram ações negociadas como concordatárias no período de 1986 a 1998. O modelo classificou corretamente 81,1% das concordatárias e 79,5% das não concordatárias.

Por fim, Minussi, Damacena e Ness Junior (2002) construíram um modelo de previsão de insolvência com a utilização da regressão logística, a partir de uma amostra constituída por 323 empresas. A precisão do modelo foi de 94,85%.

3. METODOLOGIA

Para o desenvolvimento deste estudo, foram utilizados como procedimentos os métodos de pesquisa bibliográfica e documental. A pesquisa bibliográfica é realizada a partir do levantamento de referências teóricas publicadas por meios escritos e eletrônicos, como livros e artigos científicos. Este tipo de pesquisa permite conhecer o que já se estudou sobre o assunto (FONSECA, 2002).

Por outro lado, a pesquisa documental, de acordo com Fachin (2006, p. 146) “Consiste na coleta, classificação, seleção e utilização de toda a espécie de informação. Compreende técnicas e métodos que facilitam a identificação”.

Assim, a pesquisa procede com uma abordagem quantitativa e sendo classificada quanto aos objetivos como um estudo exploratório. A pesquisa quantitativa, de acordo Gressler (2004), é caracterizada pelo desenvolvimento de hipóteses, definições de variáveis e quantificação nas modalidades de coleta de dados e informações, utilizando métodos estatísticos.

Gil (2010, p.27), afirma que “as pesquisas exploratórias têm como propósito proporcionar maior familiaridade com o problema, com vistas a torna-lo mais explícito ou a construir hipóteses. Seu planejamento tende a ser bastante flexível, pois interessa considerar os mais variados aspectos relativos ao fato ou fenômeno estudado”.

A população do estudo, a partir da qual a amostra foi selecionada, compreende as companhias abertas brasileiras que constam no cadastro da Comissão de Valores Mobiliários (CVM) em situação de recuperação judicial ou extrajudicial, contemplando 32 empresas.

Para seleção da amostra, o primeiro procedimento adotado foi a identificação das empresas que entraram com pedido de recuperação judicial entre os anos de 2011 a 2016, que possuem registro ativo na CVM e que não constava estado de falência no ano anterior ao pedido de recuperação judicial, ficando com 19 companhias no estado de recuperação judicial ou extrajudicial.

A partir da identificação das empresas insolventes, procedeu-se a seleção das empresas solventes que completariam a amostra. O grupo de solventes foi constituído selecionando-se para cada empresa insolvente, uma solvete de setor econômico e região igual.

Para aplicar os modelos de previsão de risco de crédito, foram analisadas as demonstrações financeiras de um ano antes do pedido de recuperação judicial das empresas selecionadas, abrangendo o período de 2010 a 2015. Desta forma, a amostra final utilizada no desenvolvimento deste estudo compreende 38 empresas, conforme Tabela 1.

Tabela 1. Empresas que compõem a amostra da pesquisa.

Empresas Insolventes	Empresas Solventes	Ano Evento	Ano Analisado
Cia Industrial Schlosser S.A.	Karsten S.A.	2011	2010
Buettner S.A. Ind. e Comercio.	Cia Hering.	2011	2010
Centrais Elétricas do Pará S.A	Investco S.A.	2012	2011
Tecnosolo S.A.	CR2 Empreendimentos Imobiliários S.A.	2012	2011
Teka Tecelagem Kuehnrich S.A.	Dohler S.A.	2012	2011
Refinaria Pet Manguinhos S.A.	Petro Rio S.A.	2013	2012
GPC Participações S.A.	Elekeiroz S.A.	2013	2012
Mangels Industrial S.A.	Kepler Weber S.A.	2013	2012
Óleo e Gás Participações S.A.	Qgep Participações S.A.	2013	2012
OGX Petróleo e Gás S.A.	Petróleo Brasileiro S.A. – Petrobras.	2014	2013
Inepar Equipamentos e Montagens S.A.	Whirlpool S.A.	2014	2013
Inepar S.A Ind. e Construções.	Indústrias Romi S.A.	2014	2013
Fibam Cia Industrial.	Cia Siderurgica Nacional.	2014	2013
Eneva S.A.	Itapebi Geração de Energia S.A.	2015	2014
Const Sultepa S.A.	Unicasa Indústria de Móveis S.A.	2015	2014
Lupatech S.A.	Cobrasma S.A.	2015	2014

Wetzel S.A.	Tupy S.A.	2016	2015
Oi S.A.	Telefônica Brasil S.A.	2016	2015
Viver Incorporadora e Construtora S.A.	Cyrela Brazil Realty S.A. Emprs. e Parts.	2016	2015

Partindo dessa amostra vamos utilizar vários autores já realizaram modelos de risco de crédito, sendo assim vamos utilizar as mesmas variáveis, ajustar novos modelos baseado em nossos dados e assim comparar os poderes preditivos.

Dessa forma, os modelos foram aplicados à amostra de validação com o auxílio do software *SPSS Statistics* em sua versão 20.0, sendo assim, foram realizados os testes de qualidade e significância dos modelos de Altman, Baidya e Dias; Matias; Brito e Assaf Neto; Scarpel e Milioni e construção da curva ROC para validação.

Os testes de qualidade dos modelos de análise discriminante foram realizados através do *eigenvalue* que mede a diferença entre os grupos, e a correlação canônica que mensura a associação entre as variáveis independentes e a dependente. O teste de significância foi efetuado por meio do *Lambda de Wilks* e o nível de significância foi estabelecido de acordo com a escala de Fisher, que considera os valores abaixo de 0,05 como um nível muito forte (GUIMARÃES, 2006).

Dessa forma, para os modelos de regressão logística, os testes de qualidade e significância foram efetuados por meio do *Likelihood Value* que mensura a capacidade de prever a ocorrência de um evento, o R^2 de Cox & Snell e Nagelkerke que indicam o quanto que a variável dependente é explicada pelas variáveis independentes e o teste de Hosmer e Lemeshow que verifica as diferenças entre os grupos (FERREIRA, 2010).

Por fim, a técnica estatística utilizada para validação dos modelos foi a curva ROC (*Receiver Operating Characteristic*) que está baseada nos conceitos de especificidade e sensibilidade (BARTH, 2002).

4. ANÁLISE DOS DADOS

Após a seleção da amostra de empresas e o cálculo dos índices econômico-financeiros, procedeu-se ao cálculo dos modelos no software *SPSS Statistics*, no qual utilizou-se valor 0 para indicar o estado de solvência e 1 para o estado de insolvência.

Para analisar a qualidade da função discriminante, foram calculados os autovalores das funções estimadas. De acordo com Nóbrega (2010), o autovalor demonstra a diferença entre os grupos nas funções discriminantes, dessa forma, quanto mais afastados de 1, maiores serão as variações entre os grupos explicados pela função discriminante. Para esta análise, o resultado é de apenas uma função, correspondendo a 100% da variância explicada em termos de diferenças entre grupos.

O *SPSS Statistics* demonstra a correlação canônica junto com a análise do autovalor, assim, Maroco (2007) explica que a correlação canônica mede a relação entre o grupo de variáveis que identificam os grupos e as funções discriminantes. Nesse sentido, o percentual que explica a variação entre as variáveis dependentes e independentes é determinado elevando-se ao quadrado a correlação canônica, este valor também é conhecido como R^2 (PROTÁSIO et al., 2012).

Tabela 2 - *Eigenvalues* – Altman, Baidya e Dias.

Eigenvalue	% Variância	Cumulativo %	Correlação Canônica	R^2
0,171 ^a	100	100	0,382	0,146

Fonte: Dados da pesquisa e Autores.

Tabela 3 - *Eigenvalues* - Matias.

Eigenvalue	% Variância	Cumulativo %	Correlação Canônica	R²
0,450 ^a	100	100	0,557	0,310

Fonte: Dados da pesquisa e Autores.

Conforme a tabela 2, o modelo de Altman, Baidya e Dias apresentou um autovalor de 0,171, enquanto o modelo desenvolvido por Matias, demonstrado na tabela 3, foi de 0,450, dessa forma, pode-se perceber que o segundo modelo é o que mais se aproxima de 1, demonstrando a menor variação entre os grupos. Dessa forma, o modelo de Matias apresentou maior relação entre a função discriminante entre os grupos, apresentando uma correlação canônica de 0,557, elevando este valor ao quadrado, pode-se verificar que 31% da variável dependente é explicada pelo modelo, logo, o modelo de Altman, Baidya e Dias evidenciou uma baixa correlação canônica de 0,382 com R² de 14,6%.

A significância da função discriminante é analisada por meio do teste de *Lambda de Wilk's*, que demonstra o quanto a função separa as observações em grupos diferentes. O objetivo do teste é ignorar a hipótese nula, sendo p-valor igual a zero, sendo assim, foi utilizado um nível de confiança de 95%, e o nível de significância deve ser igual ou menor que 0,05 (NÓBREGA, 2010).

Tabela 4 - Wilk's Lambda – Altman, Baidya e Dias.

Wilk's Lambda	Sig.
0,854	0,251

Fonte: Dados da pesquisa e Autores.

Tabela 5 - Wilk's Lambda – Matias.

Wilk's Lambda	Sig.
0,690	0,056

Fonte: Dados da pesquisa e Autores.

A tabela 4 demonstra os resultados do teste para o modelo de Altman, Baidya e Dias, no qual apresenta significância de 0,251 e 0,056 para a de Matias, conforme tabela 5. Dessa forma, observa-se que nenhuma das duas funções é significante na diferenciação entre os grupos.

Para avaliar a capacidade de previsão dos modelos de análise discriminante, construiu-se uma matriz de classificação para demonstrar a quantidade de empresas classificadas correta e incorretamente pelos modelos. Conforme demonstra a tabela 6, o nível de acerto do modelo desenvolvido por Altman, Baidya e Dias foi de 76,3%, tendo sido classificadas corretamente 29 empresas, por outro lado, de acordo com a tabela 7, o modelo de Matias classificou corretamente 81,6% das empresas, sendo 31 das 38 empresas da amostra.

Tabela 6 - Matriz de classificação – Altman, Baidya e Dias.

Classificação	Grupo Previsto			Total
	Solvente	Insolvente		
Grupo Original	Solvente	16	3	19
	Insolvente	6	13	19
	Solvente	84,2%	15,8%	100%
	Insolvente	31,6%	68,4%	100%

a. 76,3% dos casos originais classificados corretamente.

Fonte: Dados da pesquisa e Autores.

Tabela 7 - Matriz de classificação – Matias.

Classificação		Grupo Previsto		Total
		Solvente	Insolvente	
Grupo Original	Solvente	16	3	19
	Insolvente	4	15	19
	Solvente	84,2%	15,8%	100%
	Insolvente	21,1%	78,9%	100%

a. 81,6% dos casos originais classificados corretamente.

Fonte: Dados da pesquisa e Autores.

A análise da qualidade dos modelos de regressão logística foi realizada por meio do *Likelihood Value*, dos pseudos R^2 e do teste de Hosmer e Lemeshow. O *Likelihood Value* é um indicador da qualidade do ajustamento aos dados, quanto menor o seu valor, melhor é o ajustamento. Para medir o poder de explicação do modelo, são utilizados os Pseudos R^2 de Cox e Snell e Nagelkerke, que demonstram os efeitos que as variáveis independentes causam na variável dependente (MAROCO, 2007; ZANINI, 2007)

Para este estudo será utilizado o pseudo R^2 de Nagelkerke, que se assemelha ao coeficiente de determinação da regressão linear, sendo que quanto maior for esta medida, melhor o ajuste do modelo. Cabe ressaltar, que o pseudo R^2 de Nagelkerke tem escala de zero a um, por outro lado, o pseudo R^2 de Cox e Snell é limitado e não alcança o valor máximo de um (BRITO; ASSAF; CORRAR, 2009).

Tabela 8 - Resumo do modelo – Brito e Assaf Neto.

-2 Log Likelihood	Cox & Snell R^2	Nagelkerke R^2
27,749 ^a	0,481	0,641

Fonte: Dados da pesquisa e Autores.

Tabela 9 - Resumo do modelo – Scarpel e Milioni.

-2 Log Likelihood	Cox & Snell R^2	Nagelkerke R^2
48,035 ^a	0,115	0,153

Fonte: Dados da pesquisa e Autores.

Observa-se na tabela 8, que o -2 *Log de Likelihood* do modelo de Brito e Assaf Neto é 27, 749, enquanto o modelo de Scarpel e Milioni é 48,035, conforme demonstra a tabela 9. Assim, pode-se observar que as variáveis independentes do modelo de Brito e Assaf Neto explicam melhor a variável dependente, uma vez que é o mais próximo de zero. Da mesma forma, pode-se ainda dizer que o modelo de Brito e Assaf Neto possui um melhor ajustamento, visto que o R^2 de Nagelkerke é maior do que o apresentado no modelo de Scarpel e Milioni.

Em seguida foi realizado o teste de Hosmer e Lemeshow, que estima a relação entre os valores reais e os previstos da variável dependente. Diante disso, uma menor a diferença entre a classificação observada e a prevista, indica um melhor ajuste do modelo, portanto o

valor de qui-quadrado não deve ser significativo, ou seja, é necessário que ele seja maior do que 0,05 (HAIR et al., 2009).

Tabela 10 - Teste de Hosmer e Lemeshow – Brito e Assaf Neto.

Qui-quadrado	df	Sig.
5,158	8	0,741

Fonte: Dados da pesquisa e Autores.

Tabela 11 - Teste de Hosmer e Lemeshow – Scarpel e Milioni.

Qui-quadrado	df	Sig.
12,926	8	0,114

Fonte: Dados da pesquisa e Autores.

Conforme demonstra a tabela 10, a significância do modelo de Brito e Assaf Neto é 0,741, o que significa que o modelo é capaz de produzir 74% (setenta e quatro) por cento da estimativa, próximo ao 100% (cem) por cento, assim suas e classificações são confiáveis. Em contrapartida, o modelo de Scarpel e Milioni apresenta uma significância de 0,114, de acordo com a tabela 11, estando acima do esperado, porém em um comparativo dos dois modelos, o primeiro apresenta resultados melhores do que o segundo.

Para análise dos modelos de regressão logística, também foram desenvolvidas as matrizes de classificação. Conforme demonstra a tabela 12, o modelo de Brito e Assaf Neto classificou corretamente 78,9% das empresas, totalizando 30 empresas. Já o nível de acerto do modelo de Scarpel e Milioni, de acordo com a tabela 13, foi de 73,7%, sendo 28 das 38 empresas da amostra classificadas corretamente.

Tabela 12 - Matriz de classificação – Brito e Assaf Neto.

Classificação		Grupo Previsto		Total
		Solvente	Insolvente	
Grupo Original	Solvente	15	4	19
	Insolvente	4	15	19
Original	Solvente	78,9%	21,1%	100%
	Insolvente	21,1%	78,9%	100%

a. 78,9% dos casos originais classificados corretamente.

Fonte: Dados da pesquisa e Autores.

Tabela 13 - Matriz de classificação – Scarpel e Milioni.

Classificação		Grupo Previsto		Total
		Solvente	Insolvente	
Grupo Original	Solvente	16	3	19
	Insolvente	7	12	19
Original	Solvente	84,2%	15,8%	100%
	Insolvente	36,8%	63,2%	100%

a. 73,7% dos casos originais classificados corretamente.

Fonte: Dados da pesquisa e Autores.

Por fim, para medir a capacidade de discriminação de todos os modelos, foi desenvolvido a curva ROC. Hosmer e Lemeshow (2000, apud CABRAL, 2013), apresentam uma avaliação geral do resultado da área sob a curva ROC: área do intervalo entre 0,7 e 0,8 apresentam uma discriminação aceitável, já uma área entre 0,8 e 0,9 significa uma excelente discriminação, por fim, a área acima de 0,9 representa uma excepcional discriminação.

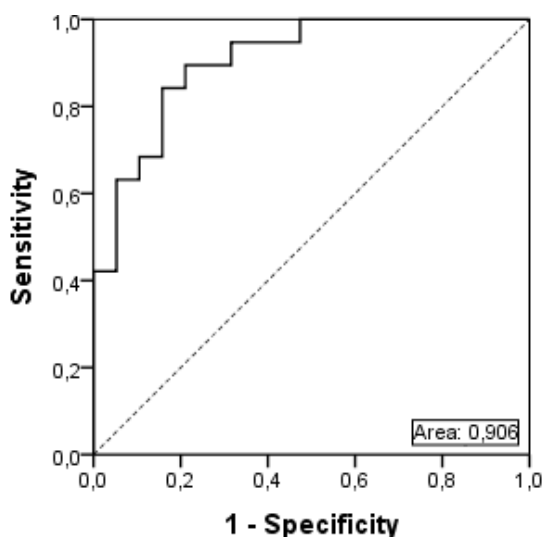


Figura 1. Curva ROC – Altman, Baidya e Dias.
Fonte: Dados da pesquisa e Autores.

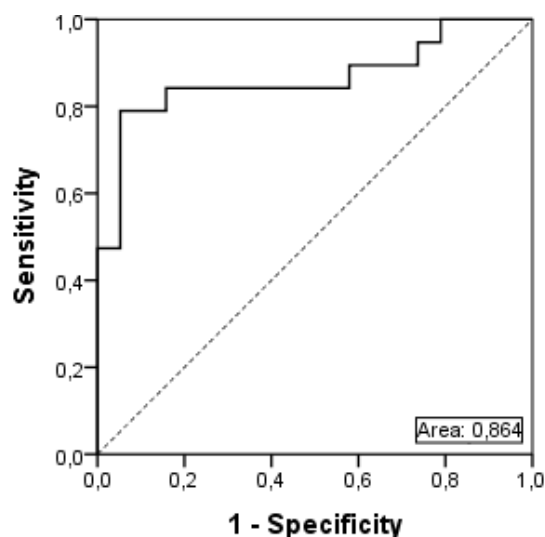


Figura 2. Curva ROC – Matias.
Fonte: Dados da pesquisa e Autores.

Observa-se na figura 1, que o modelo de Altman, Baidya e Dias obteve uma discriminação excepcional, uma vez que a área sob a curva foi de 0,906, enquanto o modelo de Matias, demonstrado na figura 2, a área sob a curva foi de 0,864 representando uma discriminação excelente.

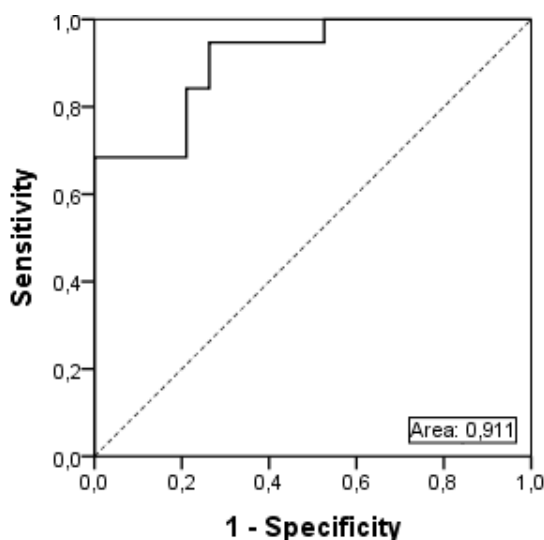


Figura 3. Curva ROC – Brito e Assaf Neto.
Fonte: Dados da pesquisa e Autores.

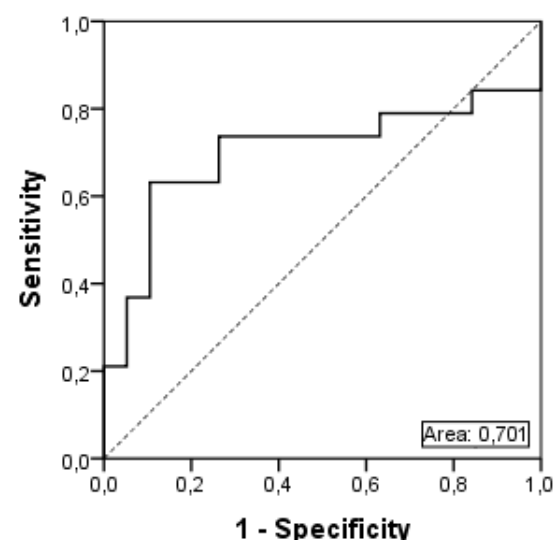


Figura 4. Curva ROC – Scarpel e Milioni.
Fonte: Dados da pesquisa e Autores.

Por fim, a curva ROC do modelo desenvolvido por Brito e Assaf Neto, representada na figura 3, revela que a área sob a curva é de 0,911, indicando um excepcional poder de discriminação do modelo. Por outro lado, o modelo de Scarpel e Milioni na figura 4 demonstram uma discriminação aceitável, visto que a área sob a curva foi de 0,701.

5. CONCLUSÃO

O objetivo deste estudo foi verificar o modelo mais adequado para classificação de risco de crédito das companhias abertas brasileiras, comparando modelos de análise discriminante e regressão logística. Para tanto, utilizou-se dois modelos para cada técnica estatística escolhida.

Os modelos de risco de crédito desenvolvidos por meio da análise discriminante apresentaram um autovalor satisfatório, no entanto a correlação canônica e o R^2 ficaram abaixo do esperado, uma vez que 31% da variável dependente é explicada pelo modelo de Altman, Baidya e Dias, e apenas 15% é explicado pelo modelo de Matias.

No que se refere à significância das funções discriminantes, pode-se observar que nenhuma das funções é significativa na diferenciação entre grupos, visto que todas apresentaram valores maiores do que 0,05. De modo geral, os modelos demonstraram bons índices de classificação, dado que o modelo de Altman, Baidya e Dias obteve 76,3% de acerto, e o modelo de Matias atingiu 81,6% de classificações corretas.

Seguindo o processo, foi realizada a análise da qualidade dos modelos de regressão logística, desta forma, pode-se constatar que o modelo desenvolvido por Brito e Assaf Neto obteve um melhor ajustamento, uma vez que apresentou um menor valor de $-2 \text{ Log de Likelihood}$ e maior pseudo R^2 . Os resultados do modelo de Scarpel e Milioni ficaram muito abaixo do esperado.

Em geral, os modelos de regressão logística apresentaram resultados satisfatórios para o teste de Hosmer e Lemeshow, no entanto, o modelo de Brito e Assaf Neto demonstrou maior capacidade de produzir estimativas e classificações confiáveis. Pode-se confirmar essa observação através da matriz de classificação, na qual o modelo de Brito e Assaf Neto obteve 78,9% de acerto, e o modelo de Scarpel e Milioni classificou corretamente 73,7% das empresas.

Sendo assim, conclui-se que a técnica de regressão logística apresenta um ótimo poder de discriminação, visto que o modelo de Brito e Assaf Neto apresentou resultados excelentes em todas as análises, além disso, esta técnica possui uma maior flexibilidade, uma vez que é possível utilizar variáveis de natureza numéricas e não numéricas.

Verifica-se que o modelo de Scarpel e Milioni apresentou resultados abaixo do esperado pois foram utilizadas apenas três variáveis independentes, que por sua vez não possuíam um bom nível de significância e qualidade. No estudo realizado pelos autores, o modelo desenvolvido alcançou 87% de classificações corretas, no entanto quando aplicado em outra amostra, não obteve o mesmo nível.

Por fim, como recomendações para pesquisas futuras, propõe-se realizar o comparativo com outras técnicas multivariadas como as redes neurais, também aplicar uma quantidade maior de modelos de previsão de insolvência e utilizar o método de Lachembruch para validação dos modelos.

REFERÊNCIAS

ALTMAN, Edward I.; BAIDYA, Tara K. N.; DIAS, Luiz Manoel Ribeiro. Previsão de problemas financeiros em empresas. **Revista de administração de empresas**, Rio de Janeiro, v. 19, n. 1, p. 17-28, jan./mar. 1979. Disponível em: <http://www.scielo.br/scielo.php?pid=S0034-75901979000100002&script=sci_arttext>.

ANDRADE, Fabio Wendling Muniz. **Desenvolvimento de modelo de risco de portfólio para carteiras de crédito a pessoas físicas**. 2004. 197 f. Tese (Doutorado em Administração de Empresas) – Fundação Getúlio Vargas, São Paulo, 2004. Disponível em: <<https://bibliotecadigital.fgv.br/dspace/handle/10438/2513>>.

BARTH, Nelson Lerner. **Métodos de discriminação entre grupos: aplicação ao problema da concessão de crédito**. 163 f. Dissertação (Mestrado em Administração) – Fundação Getúlio Vargas, São Paulo, 2002. Disponível em: <<https://bibliotecadigital.fgv.br/dspace/handle/10438/4998>>.

BESSIS, Joel. **Risk management in banking**. 3. ed. John Wiley & Sons, 2011.

BIROLO, Paula Bez; CITTADIN, Andréia; RITTA, Cleyton de Oliveira. Análise de crédito por meio de modelos de previsão de insolvência: um estudo de caso na Empresa Cerâmica Alfa SA. **Revista Catarinense da Ciência Contábil**, Florianópolis, v. 10, n. 29, p. 27-39, abr./jul. 2011. Disponível em: <<https://dialnet.unirioja.es/servlet/articulo?codigo=5454869>>.

BRITO, Giovani Antonio Silva; ASSAF NETO, Alexandre. Modelo de classificação de risco de crédito de empresas. **Revista Contabilidade & Finanças**, São Paulo, v. 19, n. 46, p. 18-29, jan.-abr. 2008. Disponível em: <<https://www.revistas.usp.br/rcf/article/view/34249>>.

BRITO, Giovani Antonio Silva; ASSAF NETO, Alexandre; CORRAR, Luiz João. Sistema de classificação de risco de crédito: uma aplicação a companhias abertas no Brasil. **Revista Contabilidade & Finanças**, São Paulo, v. 20, n. 51, p. 28-43, set./dez. 2009. Disponível em: <<https://www.revistas.usp.br/rcf/article/view/34299>>.

CABRAL, Cleidy Isolete Silva. **Aplicação do modelo de regressão logística num estudo de mercado**. 2013. 59 f. Dissertação (Mestrado em Matemática Aplicada à Economia e à Gestão) – Universidade de Lisboa, 2013. Disponível em: <<http://repositorio.ul.pt/handle/10451/10671>>.

CAOQUETTE, John B.; ALTMAN, Edward I.; NARAYANAN, Paul. **Managing credit risk: the next great financial challenge**. New York: John Wiley & Sons, 1998.

CHAIA, Alexandre Jorge. **Modelos de gestão do risco de crédito e sua aplicabilidade ao mercado brasileiro**. 121 f. Dissertação (Mestrado) – Universidade de São Paulo, São Paulo, 2003. Disponível em: <www.teses.usp.br/teses/disponiveis/12/12139/tde-21012004-224716/en.php>.

DAVIS, Karron T. Credit Scoring Software: A Brief Look. **Business Credit**, v. 101, n. 2, p. XX, fev. 1999. Disponível em: <<https://www.questia.com/read/1G1-54112217/credit-scoring-software-a-brief-look>>. Acesso em: 11 ago. 2017.

FACHIN, Odilia. **Fundamentos de metodologia**. 5. ed. São Paulo: Saraiva, 2006.

FERREIRA, Fabrício Silva. **Aplicação da técnica de regressão logística na previsão de insolvência no segmento das cooperativas de crédito no Brasil**: uma abordagem utilizando o Índice da Basiléia. 2010. 121 f. Dissertação (Administração) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010. Disponível em: <http://www.scielo.br/scielo.php?pid=S0103-20032010000400007&script=sci_arttext&tlng=es>.

FONSECA, José João Saraiva. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

GRESSLER, Lori Alice. **Introdução à pesquisa**: projetos e relatórios. 2. ed. São Paulo: Loyola, 2004.

GUIMARÃES, Ailton. **Previsão de insolvência**: um modelo baseado em índices contábeis com a utilização da análise discriminante. 2006. 100 f. Dissertação (Mestrado em Economia de Empresas) – Universidade Católica de Brasília, Brasília, 2006. Disponível em: <<https://bdtd.ucb.br:8443/jspui/handle/123456789/559>>.

HAIR, Joseph F; BLACK, William C; BABIN, Barry J; ANDERSON, Rolph E; TATHAM, Ronald L. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

MÁRIO, Pueri do Carmo. **Contribuição ao estudo da solvência empresarial**: Uma análise de modelos de previsão – estudo exploratório aplicado em empresas mineiras. 2002. 170 f. Dissertação (Mestrado em Controladoria e Contabilidade) – Universidade de São Paulo, São Paulo, 2002. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/12/12136/tde-06012006-152208/en.php>>.

MAROCO, João. **Análise estatística com utilização do SPSS**. 3. ed. Lisboa: Sílabo, 2007.

MARTINS, Márcio Severo. **A previsão de insolvência pelo modelo de Cox**: uma contribuição para a análise de companhias abertas brasileiras. 2003. 103 f. Dissertação (Mestrado em Administração) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003. Disponível em: <<https://www.lume.ufrgs.br/handle/10183/1609>>.

MATIAS, Alberto Borges. **Contribuição às técnicas de análise financeira**: um modelo de concessão de crédito. 1978. 101 f. Dissertação (Mestrado) – Universidade de São Paulo, São Paulo, 1978.

MENDES, Ivan Oliveira de Vieira. **Variáveis discriminantes dos estágios de insolvência de empresas**. 2014. 111 f. Dissertação (Mestrado em Contabilidade) – Universidade Federal do Paraná, Curitiba, 2014. Disponível em: <<https://acervodigital.ufpr.br/handle/1884/35681>>.

MINUSSI, João Alberto; DAMACENA, Cláudio; NESS JUNIOR, Walter Lee. Um modelo de previsão de solvência utilizando regressão logística. **Revista de Administração Contemporânea – RAC**, Curitiba, v. 6, n. 3, p. 109-128, set.-dez. 2002. Disponível em: <http://www.scielo.br/scielo.php?pid=S1415-65552002000300007&script=sci_arttext&tlng=es>.

NÓBREGA, Diogo Medeiros. **Análise discriminante utilizando o software SPSS**. 2010. 54 f. Trabalho de Conclusão de Curso (Bacharelado) – Universidade Estadual da Paraíba, Campina Grande, 2010. Disponível em: <<http://dspace.bc.uepb.edu.br/jspui/handle/123456789/498>>.

PINHEIRO, Laura Edith Taboada et al. Validação de modelos brasileiro de previsão de insolvência. **Contabilidade Vista & Revista**. Minas Gerais, v. 18, n. 4, p. 83-103, out./dez. 2007. Disponível em: <<https://revistas.face.ufmg.br/index.php/contabilidadevistaerevista/article/view/341>>.

PITA, Fabiano Veraldo da Costa; ALBUQUERQUE, Lucia Galvão de. Efeitos da utilização de diferentes covariáveis na avaliação do ganho de peso médio diário em suínos. **Revista Brasileira de Zootecnia**, p. 736-743, 2001. Disponível em: <<https://repositorio.unesp.br/handle/11449/4843>>.

PROTÁSIO, Thiago de Paula et al. Análise de correlação canônica entre características da madeira e do carvão vegetal de *Eucalyptus*. **Scientia Forestalis**. Piracicaba, v. 40, n. 95, p. 317-326, set. 2012. Disponível em: <https://www.researchgate.net/profile/Paulo_Trugilho/publication/288569533_Canonical_correlation_analysis_between_characteristics_of_Eucalyptus_wood_and_charcoal/links/5698e40b08aea2d74377242c/Canonical-correlation-analysis-between-characteristics-of-Eucalyptus-wood-and-charcoal.pdf>.

SANVICENTE, Antonio Zorato; MINARDI, Andrea Maria A. Fonseca. Identificação de indicadores contábeis significativos para a previsão de concordata de empresas. **Instituto Brasileiro de Mercado de Capitais**, Working Paper, 1998. Disponível em: <https://www.researchgate.net/profile/Andrea_Minardi/publication/5104691_Identificacao_de_indicadores_contabeis_significativos_para_previsao_de_concordata_de_empresas/links/59763794aca2728d02705f23/Identificacao-de-indicadores-contabeis-significativos-para-previsao-de-concordata-de-empresas.pdf>.

SCARPEL, Rodrigo Arnaldo; MILIONI, Armando Zeferino. Utilização conjunta de modelagem econométrica e otimização em decisões de concessão de crédito. **Pesquisa Operacional**, São Paulo, v. 22, n. 1, p. 61-72, jan./jun. 2002. Disponível em: <http://www.scielo.br/scielo.php?pid=S0101-74382002000100004&script=sci_arttext&tlng=pt>.

SERASA EXPERIAN, **Comércio, indústria e serviços batem recorde de pedidos de recuperação judicial em 2016. 2017**. Disponível em <<http://noticias.serasaexperian.com.br/blog/2017/01/12/comercio-industria-e-servicos-batem-recorde-de-pedidos-de-recuperacao-judicial-em-2016-mostra-serasa-experian/>>. Acesso em: 04 ago. 2017.

SILVA, José Pereira da. **Modelos para classificação de empresas com vistas à concessão de crédito**. 1982. Dissertação (Mestrado) - Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo. Disponível em: <<https://bibliotecadigital.fgv.br/dspace/handle/10438/10981>>.

STÜPP, Diego Rafael. **Previsão de insolvência a partir de indicadores contábeis: evidências de empresas listadas na BM&FBOVESPA nos anos de 2004-2013**. 2015. 119 f. Dissertação (Mestrado em Contabilidade) – Universidade Federal de Santa Catarina, Florianópolis, 2015. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/134789>>.

ZANINI, Alexandre. **Regressão logística e redes neurais artificiais: um problema de estrutura de preferência do consumidor e classificação de perfis de consumo**. 2007. Dissertação (Mestrado em Economia Aplicada) – Universidade Federal de Juiz de Fora, Juiz de Fora, 2007. Disponível em: <http://www.ufjf.br/poseconomia/files/2010/01/td_007_20071.pdf>.