

Avaliação do Desempenho de Técnicas de Classificação Aplicadas à Previsão de Insolvência de Empresas de Capital Aberto Brasileiras

Rômulo Alves Soares – Mestrando
Universidade Federal do Ceará
romuloalves61@gmail.com

Sílvia Maria Dias Pedro Rebouças – Doutora
Universidade Federal do Ceará
smdpdro@gmail.com

Resumo

O presente estudo tem como objetivo a avaliação do desempenho de técnicas de classificação aplicadas à previsão de insolvência de empresas brasileiras de capital aberto. Para esse fim, foram utilizadas informações contábeis de empresas brasileiras de capital aberto, compondo uma amostra com 87 companhias, sendo 21 delas insolventes e 66 solventes, compreendendo um período de análise de 2002 a 2012. A comparação dos modelos foi feita por meio da área abaixo da curva ROC, bem como a taxa de precisão (*accuracy ratio*). Para a execução do objetivo geral, buscou-se, secundariamente, aplicar técnicas de classificação para a construção de modelos de classificação para empresas, e também observar quais indicadores foram os mais importantes para as análises. As técnicas utilizadas foram Análise Discriminante Linear, Regressão Logística, Vizinhos Mais Próximos, Árvores de Classificação e Redes Neurais Artificiais. Os resultados obtidos para as cinco técnicas empregadas mostram um bom desempenho para todos. O modelo com melhor performance foi o de Redes Neurais Artificiais, enquanto o com pior desempenho foi o Classificador dos Vizinhos mais Próximos. O indicador mais importantes foi a relação entre o Capital Circulante Líquido e o Ativo Total, considerado importante em três modelos, seguido do Endividamento Geral, e da Margem Líquida, relevantes para dois modelos apenas, ressaltando que esse tipo de análise apenas foi possível em três das cinco técnicas empregadas.

Palavras-chave: Previsão de insolvência. Técnicas de Classificação. Avaliação de desempenho.

Área Temática: Finanças Corporativas e Mercado Financeiro.

1 Introdução

A previsão de insolvência constitui uma ferramenta importante para o desenvolvimento econômico. Gestores de uma firma, por exemplo, podem usar esse tipo de informação para mudar decisões acerca do futuro de sua empresa, já que o processo de deterioração financeira ocorre, em geral, de maneira progressiva.

Uma das aplicações mais conhecidas da previsão de insolvência é auxiliar na redução do risco de crédito. Segundo Caouette *et al* (2008), sempre que um indivíduo faz uso de um produto ou serviço sem que haja o pagamento imediato pelo mesmo, é possível identificar elementos relacionados ao risco de crédito. Quando uma compra a prazo é realizada, a empresa que cede o produto está sendo exposta a um risco de não receber o valor devido. Ou ainda, empresas que fornecem serviços básicos de maneira contínua, como distribuição de água e energia elétrica, também se expõem ao risco de crédito, já que o pagamento pelos serviços prestados só é efetuado ao final de um certo período, normalmente mensal. Assim, o risco de crédito é, em outras palavras, a possibilidade de que um credor sofra uma perda em razão do não cumprimento de obrigações assumidas por terceiros.

No mercado financeiro, muitas transações envolvem a disponibilidade de recursos para tomadores, seja em forma de empréstimo ou financiamento, mediante o comprometimento de pagamento em uma data posterior. Com isso, é natural que bancos e outras instituições que forneçam crédito, ou mesmo investidores, busquem formas de otimizar a utilidade dos recursos repassados às empresas tomadoras de recursos, reduzindo ao máximo o risco de crédito.

Nesse sentido, a capacidade de prever insolvência tem um papel fundamental. Do ponto de vista econômico, diz-se que uma companhia está insolvente quando o total de seu passivo é superior ao seu ativo, ou seja, quando ela não pode pagar todas as dívidas assumidas mesmo com a liquidez total de seus bens e direitos. Por isso, estabelecer com antecedência quais empresas apresentam propensão à insolvência e quais são financeiramente saudáveis, é de vital importância para credores e investidores. Para ilustrar isso, analisa-se o caso das Lojas Arapuã, grupo empresarial que durante a década de 90 chegou a ser o maior varejista de eletrodomésticos no Brasil. Foi eleita na edição de julho de 1997 da revista Exame Melhores e Maiores como a melhor empresa de comércio varejista do ano anterior. No entanto, já em 1998 as Lojas Arapuã viram-se forçadas a pedir concordata por conta de dificuldades financeiras, o que gerou um imbróglio judicial entre a empresa e parte de seus credores e fornecedores, ainda não resolvido até hoje.

A utilização de técnicas de análise multivariada, trazidas para a problemática da previsão de insolvência por Altman (1968), é a principal corrente metodológica no que diz respeito a esse campo de estudos. Diversas técnicas de classificação têm sido empregadas com o intuito de construir modelos capazes de prever com eficiência a entrada em estado de insolvência de empresas.

Engelmann, Hayden e Tasche (2003), baseados em orientações emitidas pelo *Basel Committee on Banking Supervision* em 2000 e 2001, afirmam que a modelagem de risco de crédito muito se desenvolveu com o passar dos anos, e que o desenvolvimento de estatísticas confiáveis para a avaliação desses modelos adquire importância, uma vez que modelos de baixa qualidade podem levar a uma alocação de capital subótima.

Assim, este trabalho apresenta como questão de pesquisa: qual o nível de desempenho das técnicas estatísticas mais comumente utilizadas na previsão de insolvência de empresas brasileiras de capital aberto com base em indicadores contábeis?

Com base na literatura pesquisada, formula-se a hipótese de que técnicas de classificação empregadas na previsão de insolvência apresentam, em geral, bom desempenho (BRITO; ASSAF NETO, 2008; BRITO; ASSAF NETO; CORRAR, 2009; BRESSAN *et al.*, 2011; JACKSON; WOOD, 2013).

Dessa forma, o objetivo geral deste trabalho é avaliar o desempenho de técnicas estatísticas de classificação empregadas na problemática da previsão de insolvência. Especificamente, esse estudo se propõe a: i) aplicar técnicas de classificação para modelagem de insolvência de empresas brasileiras; e ii) identificar os indicadores contábeis mais importantes para a diferenciação de empresas segundo sua situação financeira.

O evento estudado foi a entrada em estado de insolvência, e o seu início foi definido como a data em que uma empresa tenha feito pedido formal de concordata ou recuperação judicial. A amostra final contou com 21 empresas insolventes, e 66 empresas solventes escolhidas conforme a distribuição setorial do primeiro grupo. O período analisado compreende os anos de 2002 a 2012. A coleta dos dados foi realizada no *software Economatica*, do qual foram obtidas informações contábeis do ano imediatamente anterior a ocorrência do evento para empresas insolventes e as demais emparelhadas. As técnicas empregadas para classificação são: análise discriminante (LDA), regressão logística (RL), classificador dos vizinhos mais próximos (kNN), árvores de classificação e regressão (CART) e redes neurais artificiais (ANN). Para a comparação das técnicas foram empregadas as curvas ROC e estatísticas dela derivadas.

O presente estudo visa contribuir com a discussão sobre o tema ao adicionar mais evidências empíricas em aspectos diferenciados da previsão de insolvência. Apresenta ainda uma metodologia pouco explorada, em especial no Brasil, para a avaliação do desempenho de modelos de risco de crédito.

2 Referencial Teórico

2.1 Insolvência de empresas

Insolvência, do ponto de vista empresarial, pode ser entendida de diversas maneiras. Chung, Tan e Holdsworth (2008) citam dois conceitos para insolvência. O primeiro deles está ligado ao fluxo de caixa e diz que uma companhia que não pode pagar suas dívidas antes do vencimento, é considerada insolvente. Esta situação, no entanto, pode ser fruto da incapacidade da empresa de realizar seus ativos antes dos vencimentos dos débitos, e não necessariamente da insuficiência deles. Isso abre margem para outra definição de insolvência dada pelos autores, ligada ao balanço, na qual uma empresa é considerada insolvente se o total do passivo exceder o total do ativo, mesmo que a companhia possa honrar os seus compromissos de curto prazo.

Segundo Ross (*apud* SILVA, 2006) o conceito de insolvência não é definido de maneira precisa. Há uma grande variedade de eventos que podem caracterizar, seja de maneira isolada ou em conjunto, o estado de insolvência em uma empresa. O mesmo autor ainda lista alguns sintomas da insolvência: i) redução de dividendos; ii) fechamento de unidades; iii) prejuízos; iv) dispensa de funcionários; v) renúncias de presidentes; vi) quedas substanciais do preço da ação.

Altman (1968) considerou insolventes em seu estudo, as empresas que fizeram pedido de concordata segundo a legislação americana. A Constituição dos Estados Unidos define como insolvente uma entidade cujo total de dívidas exceda a soma de seus direitos, considerados em valor justo, ou que não tenham capacidade de quitar as suas obrigações, na medida em que elas passem a ser exigíveis.

No Brasil, o primeiro dispositivo legal que criava instrumentos jurídicos para o enfrentamento de dificuldades financeiras de empresas foi o Decreto Lei nº 7.665/1945, a Lei de Falências e Concordatas. A concordata era definida como uma ação na qual uma empresa devedora poderia renegociar os prazos de vencimento de dívidas ou reemitir débitos, objetivando a solução do seu passivo quirografário, e consequentemente, evitar ou suspender o processo de falência, em caso de concordata preventiva ou suspensiva.

Em 2005, com o surgimento da Lei nº 11.101, Lei de Recuperação de Empresas e Falências, a concordata foi substituída pela recuperação judicial. A concordata definida pela lei anterior possuía uma conotação de favorecimento legal, uma vez que era concedida por um juiz que julgasse que a empresa devedora agia de boa-fé, independente da concordância ou não dos credores. A recuperação judicial assume um caráter contratual, já que para que possa ser efetivada é necessário que 3/5 dos credores a aprovem, o que torna o seu cumprimento obrigatório para todas as partes. A recuperação judicial só pode ser iniciada antes do processo de falência, diferente da concordata que também poderia ser iniciada durante a falência, tendo efeito suspensivo. A recuperação judicial não pode ser requerida por empresas públicas, sociedades de economia mista, instituições financeiras, cooperativas de crédito, consórcios, entidades de previdência complementar, planos assistenciais de saúde, sociedades securitárias e sociedades de capitalização (CLARO, 2008).

A nova lei também mudou o entendimento sobre falência. Antigamente a sua finalidade era fazer com que uma empresa pagasse o que era devido aos seus credores, ou o que fosse possível ser liquidado. Com a mudança, a falência passou a ser vista como o processo de retirada de empresas irrecuperáveis do mercado. Somente após a retirada da empresa é que se preocupará com a quitação das dívidas, o que passou a denominar-se judicialmente de liquidação.

Estudos de previsão de insolvência que utilizam empresas brasileiras costumemente definem como evento a entrada em concordata, recuperação judicial ou falência. Isso pode ser visto, por exemplo, em Altman, Baydia e Dias (1979) e Horta (2010).

Existem também aqueles que, por utilizarem uma base de dados de bancos ou outras instituições interessadas nesse tipo de estudo, utilizam uma classificação prévia de empresas solventes e insolventes realizadas por essas instituições, como é o caso, por exemplo, do trabalho de Elizabetsky (1976), que utilizou dados de empresas clientes do Banco Comercial. Na bibliografia pesquisada ainda encontraram-se estudos publicados no Brasil que consideram como data de entrada em insolvência o dia em que as ações de determinada empresa passaram a ser negociadas como concordatárias na Bolsa de Valores de São Paulo, como é o caso de Sanvicente e Minardi (1998).

2.2 Estudos sobre previsão de insolvência

Em 1932, Fitzpatrick escreveu o artigo *A Comparison of the Ratios of Successful Industrial Enterprises With Those of Failed Companies*, no qual comparou 13 indicadores de 20 empresas falidas, com os de 20 não-falidas, emparelhadas de acordo com o setor econômico. Chegou à conclusão de que, na maioria dos casos, empresas bem sucedidas apresentavam indicadores superiores aos de empresas que vieram a falir. Fitzpatrick (*apud* Anjum, 2012) identificou cinco estágios que antecedem a falência: i) incubação, quando os problemas financeiros começam a surgir; ii) constrangimento financeiro, estágio no qual a administração toma ciência das dificuldades da empresa; iii) insolvência financeira, quando a firma é incapaz de adquirir recursos suficientes para cobrir suas obrigações; iv) insolvência geral, que ocorre quando o total do passivo excede o ativo; e v) insolvência legal, quando há procedimentos legais que buscam proteger os credores da empresa ou quando ocorre a sua liquidação.

Em 1966, Beaver levou o estudo da previsão de insolvência para um novo estágio. Em seu trabalho *Financial Ratios as Predictors of Failure*, Beaver utilizou os dados de empresas industriais dos EUA que faliram entre os anos de 1954 e 1964, emparelhadas, por setor e tamanho do ativo, com empresas economicamente saudáveis do mesmo período. Foram considerados no estudo os cinco anos que antecederam o pedido formal de falência. A amostra inicial contava com 158 empresas, porém, devido à indisponibilidade dos dados ao longo dos cinco anos, essa quantidade foi diminuindo, chegando a 117 no quinto ano que

antecedeu a falência. Com base nas demonstrações financeiras destas empresas, Beaver montou 30 índices, divididos em seis grupos.

Beaver (1966) fez, inicialmente, uma análise de perfil e concluiu que os índices das empresas falidas se deterioraram com muito mais rapidez do que os das empresas que permaneceram saudáveis. Posteriormente, o autor testou a habilidade preditiva de cada um dos indicadores, montando modelos com cada um deles. Ao concluir o seu trabalho, Beaver sugere que estudos posteriores utilizem vários indicadores simultaneamente na construção dos modelos, o que acabaria por determinar a tendência dos trabalhos vindouros acerca da previsão de insolvência.

O primeiro estudo que utilizou alguma forma de análise multivariada para fins de previsão de insolvência foi o de Altman (1968). A técnica escolhida pelo autor foi a Análise Discriminante Linear (LDA). Naquela época havia uma predominância na área financeira da Análise de Regressão Múltipla, porém, Altman julgou a LDA mais apropriada para o estudo ao qual se propusera. Utilizou uma amostra de 66 empresas, 33 em cada grupo. As empresas consideradas insolventes no estudo foram aquelas que fizeram pedido judicial de concordata de acordo com a lei americana, entre os anos de 1946 e 1965, todas do setor industrial. As empresas solventes foram escolhidas de maneira pareada às insolventes, levando em conta o ramo de atuação, e tamanho do ativo que não extrapolasse os limites do primeiro grupo, que ficara \$ 0,7 milhão e \$ 25,9 milhões.

O modelo proposto por Altman (1968) obteve uma precisão geral de 95,45% para um ano antes do evento da concordata e 83,08% para dois anos. O autor também faz uma diferenciação entre os tipos de erro do modelo. Ele chama de Erro Tipo I aquele em que uma empresa insolvente é classificada como solvente, enquanto o Erro Tipo II é a classificação de uma empresa solvente como insolvente. A literatura sugere que o primeiro tipo de erro é mais preocupante do que o segundo, uma vez que para um credor é mais prejudicial perder o investimento feito em uma empresa que venha a quebrar, do que o custo de oportunidade de deixar de investir em uma empresa saudável (CAOINETTE *apud* CASTRO JÚNIOR, 2003). Nesse modelo, o Erro Tipo I foi de 6,06% e 28,13% e o Erro Tipo II de 3,03% e 6,06%, para um e dois anos anteriores à falência, respectivamente..

Desde o primeiro modelo proposto por Altman, muitos outros foram surgindo, e conforme a tecnologia foi avançando, novas técnicas mais poderosas puderam ser aplicadas ao problema de se prever a entrada em estado de insolvência. Atualmente é possível empregar, além das técnicas estatísticas clássicas como a análise discriminante e a regressão logística, técnicas de inteligência computacional, como é o caso das redes neurais artificiais e máquinas de suporte vetorial.

Chung, Tan e Holdsworth (2008) realizaram um trabalho no qual utilizaram os dados de 10 empresas com problemas financeiros, definido como pedido formal de concordata, detectados entre 2005 e 2007, e compararam com outras 35 empresas sem dificuldades no mesmo período, todas as empresas do setor financeiro da Nova Zelândia. Calcularam 36 indicadores utilizando balanços até três anos anteriores ao pedido de concordata. Compararam o desempenho obtido utilizando a LDA e as ANN por meio de um teste t chegando à conclusão de que o segundo modelo é mais eficiente.

Horta *et al* (2011), com o intuito de testar métodos diferentes para seleção de indicadores para serem utilizados em estudos de previsão de insolvência, utilizou árvores de classificação. A base de dados desse trabalho foi composta, inicialmente, por empresas listadas no Serasa e na Bovespa como concordatárias, em recuperação judicial ou falidas, durante o período de 2005 a 2007. Posteriormente, buscaram outras empresas saudáveis que atuassem no mesmo setor, com tamanho do ativo semelhante e, quando possível, localizadas na mesma região das empresas do primeiro grupo. A amostra final apresenta 56 empresas insolventes e 112 solventes. Para validar os resultados, foi utilizada a validação cruzada, que

consiste na divisão do conjunto original de dados em k subconjuntos menores, sendo estimado um modelo utilizando $k-1$ desses grupos, que é validado com o conjunto que ficou de fora da estimação. Horta *et al* (2011) utilizaram dez subconjuntos para a validação cruzada. Foram construídos três modelos utilizando metodologias diferentes para a seleção dos dados, com percentuais de acerto de 89,88%, 91,66% e 92,26%.

Arieshanti *et al* (2013) comparou modelos de previsão de insolvência criados a partir do método kNN com outros construídos por Redes Neurais e Máquina de Suporte Vetorial. Em seu estudo, o autor utilizou a base de dados disponibilizada por Wieslaw em seu *website*, que é composta por 112 empresas insolventes e 128 solventes, considerando um período de dois a cinco anos anterior ao pedido de concordata. Os modelos baseados em metodologias de kNN obtiveram 77,50% e 75,42% de precisão, aqueles feitos por Redes Neurais acertaram em 74,50% e 71,00%, enquanto os baseados em Máquinas de Suporte Vetorial obtiveram 71,58% e 70,42%.

Jackson e Wood (2013) utilizaram as propriedades da curva ROC para avaliar diversos modelos de classificação baseados em dados contábeis como o de Beaver (1966), Altman (1968), Taffler (1983) e o modelo logístico de Ohlson (1980), usando informações de empresas britânicas em um período posterior à adoção do IFRS no país. Também construíram um modelo próprio baseado em redes neurais artificiais, usando os mesmos indicadores. Incluíram também outros modelos baseados em informações do mercado. Os autores concluíram que a capacidade preditiva era, em geral, consideravelmente menor do que o que é reportado. Viram também que o modelo baseado em RNA apresentou o melhor desempenho dentre aqueles que utilizaram dados contábeis.

3 Metodologia

Essa pesquisa se caracteriza, de acordo com Collis e Hussey (2005), como sendo descritiva, em relação a seus objetivos, e quantitativa, quando aos procedimentos, uma vez que busca, por meio da utilização de dados secundários e técnicas de análise estatísticas, avaliar o desempenho de técnicas de classificação aplicadas ao problema de insolvência de empresas de capital aberto.

A população considerada nesse estudo são as empresas brasileiras de capital aberto listadas na BMF&BOVESPA, que estiveram em atividade entre os anos de 2003 e 2012. A escolha de empresas dessa natureza se dá pela relativa facilidade quanto à obtenção de dados para o estudo proposto, sendo empresas de domínio público as mais utilizadas para pesquisas desse tipo. A quantidade média de empresas listadas durante o período em questão foi de 372, sendo 2005 o ano com menos empresas (343), e 2007 o que apresentou a maior quantidade (404).

Para a composição da amostra, inicialmente buscou-se no sítio da CVM quais empresas haviam feito pedido de concordata preventiva, para os anos anteriores de 2003 e 2005, ou petição inicial de recuperação judicial, para os anos de 2006 a 2012. Dessa busca, obtiveram-se 22 empresas, sendo uma delas descartada por ter feito pedido de concordata preventiva em 1998, o qual foi aceito, não sendo bem sucedida, no entanto, realizando, já em 2009, pedido de recuperação judicial, aproveitando-se da mudança na legislação brasileira ocorrida em 2005.

As 21 restantes foram consideradas insolventes para esse estudo, sendo em seguida distribuídas de acordo com o segmento de atuação. Para esse fim foi utilizada a classificação do NAICS nível 2, disponível no Economatica, que conta com 92 tipos de setores. Buscou-se então empresas consideradas solventes durante o período analisado, isto é, que não tenham pedido concordata ou recuperação judicial entre 2003 e 2012 para que pudessem ser emparelhadas, de acordo com o setor, com as firmas insolventes. Para cada empresa com

dificuldades financeiras, foram encontradas até cinco saudáveis, de acordo com a disponibilidade dos dados.

Tendo em vista que algumas empresas insolventes apresentaram classificação única segundo o NAICS-2, recorreu-se ao subsetor ao qual a empresa se encontrava na listagem da BOVESPA, que contava, na época da coleta de dados, com 42 categorias. A amostra final é composta por 87 companhias, sendo 21 insolventes e 66 solventes.

Dessas firmas foram coletadas informações contábeis do ano imediatamente anterior ao pedido de concordata das empresas insolventes e das solventes com elas emparelhadas. Com base nessas informações foram calculados 16 indicadores contábeis, cuja escolha foi baseada em estudos anteriores realizados no Brasil. Pereira, Domínguez e Ocejó (2007) afirmam que a evidência empírica permite constatar que a escolha de indicadores que tiveram bons desempenho em trabalhos prévios leva, geralmente, a bons resultados. Os indicadores considerados para o estudo são apresentados no Quadro 02.

A base de dados formada com esses indicadores é analisada em quatro etapas: i) elaboração de estatísticas descritivas; ii) análise fatorial para a redução do número de indicadores; iii) aplicação de técnicas de classificação para as variáveis resultantes da análise fatorial; e iv) comparação dos resultados obtidos na etapa anterior.

Na primeira etapa foram utilizadas tabelas de frequência para aferir a distribuição de empresas por setor, enquanto para os indicadores foram calculadas medidas de tendência central, variabilidade e posição.

Na segunda parte foi aplicada uma análise fatorial, que, segundo Hair *et al.* (2005), analisa a estrutura de correlações existentes entre as variáveis e define dimensões latentes comuns, denominadas fatores. A aplicação deste método justifica-se pois não é desejável que variáveis com forte correlação com outras variáveis sejam incluídas em modelos. Esse fenômeno é conhecido como multicolinearidade, o qual influencia nos erros padrões dos coeficientes, fazendo com que sejam menores, o que dificulta a estimação dos parâmetros. Como a Análise Fatorial parte do pressuposto que variáveis altamente correlacionadas geram agrupamentos, esse método pode ser empregado para evitar esse tipo de problema. O número de fatores foi escolhido de acordo com o critério da raiz latente. De acordo com Hair *et al.* (2005) esse é o critério mais utilizado para a definição da quantidade de fatores e baseia-se no fato de que cada fator individual deve explicar pelo menos uma variável, assim a análise é feita enquanto os fatores possuem autovalores maiores do que um. A escolha das variáveis foi feita de acordo com o critério da variável substituta, isto é, dentro de cada fator foi escolhida a variável com maior poder de explicação, sendo descartadas as demais. Para verificar a adequação dos dados a essa técnica, foram utilizado o teste esfericidade de *Bartlett*, que testa a hipótese nula de que a matriz de correlação da amostra é uma identidade, o que tornaria a análise fatorial inadequada. Também foi usado o critério de Kaiser-Meyer-Olkin, que aponta qual a proporção de variância dos dados pode ser considerada comum a todas as variáveis. Segundo essa estatística, a análise fatorial é aplicável quando o valor observado do KMO é superior à 0,5.

Quadro 1 – Indicadores contábeis utilizados no estudo

Variável	Fórmula	Utilização em estudos anteriores
X ₁	(Ativo Circulante - Passivo Circulante)/Ativo Total	Altman, Baydia e Dias (1979); Sanvicente e Minardi (1998); Brito, Assaf Neto e Corrar (2009)
X ₂	LAJIR/Ativo Total	Altman, Baydia e Dias (1979); Sanvicente e Minardi (1998)
X ₃	Patrimônio Líquido/Exigível Total	Altman, Baydia e Dias (1979); Onusic <i>et al.</i> (2006); Horta (2010)
X ₄	Receita Líquida/Ativo Total	Onusic <i>et al.</i> (2006); Brito, Assaf Neto e Corrar (2009); Horta (2010)

X ₅	Exigível Total/Ativo Total	Lachtermacher e Espenchitt (2001); Brito, Assaf Neto e Corrar (2009); Horta (2010)
X ₆	Ativo Circulante/Passivo Circulante	Matias (1976); Kanitz (1978); Brito, Assaf Neto e Corrar (2009)
X ₇	Lucro Líquido/Ativo Total	Lachtermacher e Espenchitt (2001); Scarpel (2008); Horta (2010)
X ₈	(Ativo Circulante - Estoques)/Passivo Circulante	Kanitz (1978); Onusic <i>et al.</i> (2006); Horta (2010)
X ₉	(Ativo Circulante + Realizável a Longo Prazo)/Exigível Total	Kanitz (1978); Onusic <i>et al.</i> (2006); Scarpel (2008)
X ₁₀	LAJIR/Despesas Financeiras	Sanvicente e Minardi (1998); Brito, Assaf Neto e Corrar (2009); Carvalho <i>et al.</i> (2010)
X ₁₁	Lucro Líquido/Receita Líquida	Elizabetsky (1976); Minussi, Damascena e Ness Júnior (2002); Horta (2010)
X ₁₂	LAJIR/Exigível Total	Guimarães e Moreira (2008)
X ₁₃	Disponibilidades/Passivo Circulante	Brito, Assaf Neto e Corrar (2009); Aita, Zani e Silva (2010); Horta (2010)
X ₁₄	Estoques/Ativo Total	Elizabetsky (1976); Minussi, Damascena e Ness Júnior (2002); Brito, Assaf Neto e Corrar (2009)
X ₁₅	Disponibilidades/Ativo Permanente	Elizabetsky (1976); Lanchtermacher e Espenchitt (2001)
X ₁₆	Ativo Permanente/Parimônio Líquido	Lanchetermacher e Espenchitt (2001); Casro Júnior (2003); Brito, Assaf Neto e Corrar (2009)

Fonte: Dados da pesquisa.

Na terceira etapa procedeu-se com a aplicação das técnicas de classificação. Inicialmente foi verificada a hipótese de normalidade multivariada dos dados por meio do teste de *Shapiro-Wilk*, a qual foi rejeitada. Apesar disso, a técnica de análise discriminante (LDA) foi empregada nesse estudo, dada sua relevância em pesquisas de previsão de insolvência. Além da LDA, as outras técnicas utilizadas no trabalho são: regressão logística (RL), classificador dos vizinhos mais próximos (kNN), árvores de classificação e regressão (CART) e redes neurais artificiais (ANN).

O método de LDA objetiva classificar uma observação dentro de um ou mais grupos levando em consideração as suas características individuais. É indicada quando a variável dependente é qualitativa (solvente e insolvente para os estudos de previsão de insolvência), enquanto a Regressão Múltipla usa variáveis dependentes quantitativas. Para que essa técnica seja aplicada é necessário que os grupos sejam estabelecidos, para a partir daí, com base nas características de cada grupo, ou seja, as variáveis independentes (índices financeiros), a LDA extraia uma combinação linear desses fatores que seja capaz de discriminá-los dentro dos grupos previamente estabelecidos. Essa técnica surgiu, segundo Corrar *et al* (2012), em 1935, em um estudo para classificação de plantas em duas populações feito por R. A. Fisher.

A RL é uma técnica de classificação aplicável quando a variável de resposta é dicotômica. Algumas considerações teóricas e práticas sugerem que, quando a variável resposta é binária, a forma da função resposta será frequentemente sigmoideal. Para esses casos a função logística é a mais utilizada. Sua expressão matemática é dada por:

$$E[Y] = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)} = [1 + \exp(-\beta_0 - \beta_1 X_1 - \dots - \beta_p X_p)]^{-1}$$

em que Y é a variável resposta, X₁, ..., X_p são as variáveis explicativas e β_0 , ..., β_p são os coeficientes estimados a partir do conjunto de dados por meio da máxima verossimilhança.

Uma propriedade bastante útil da função logística é o fato dela poder ser linearizada. Definindo $\pi(x) = E[Y|x] = P(Y=1|x)$, pode ser aplicada a transformação *logit*, dada por:

$$\text{logit}[\pi(x)] = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right)$$

logo, a função resposta *logit* pode ser expresso como:

$$\text{logit}[\pi(x)] = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

A razão entre $\pi(x)$ e $1 - \pi(x)$ é chamada de *odds ratio*, ou razão de chance, pois representa uma relação entre as probabilidades de sucesso e fracasso.

O kNN é um dos métodos mais simples para a classificação de uma amostra. O modelo baseia-se na distância entre pares de observações. Nele há dois parâmetros a serem escolhidos: a função da distância a ser utilizada e o valor dos k vizinhos mais próximos.

No que diz respeito ao primeiro parâmetro, a distância euclidiana é a mais utilizada. Considerando $\mathbf{x}_n = (x_{n1}, x_{n2}, \dots, x_{np})$ e $\mathbf{x}_k = (x_{k1}, x_{k2}, \dots, x_{kp})$, duas observações com dimensão p , a distância euclidiana entre elas é dada por:

$$d(\mathbf{x}_n, \mathbf{x}_k) = \|\mathbf{x}_n - \mathbf{x}_k\| = \sqrt{(\mathbf{x}_n - \mathbf{x}_k)^2} = \sqrt{\sum_{i=1}^p (x_{ni} - x_{ki})^2}$$

Depois de computadas todas as distâncias de $\mathbf{x}$ para as demais observações, as k mais próximas, sendo este um número inteiro maior que ou igual a um, irão constituir a vizinhança da observação $\mathbf{x}$. A classificação do modelo para $\mathbf{x}$, então, será dada pela moda da classificação da vizinhança.

Já a CART, segundo Basgalupp (2010), foi proposta por Breiman *et al* (1984) e consiste em uma técnica não paramétrica que induz tanto árvores de classificação, caso a variável dependente seja categórica, quanto árvores de regressão, caso a variável dependente seja contínua. Ainda segundo o autor, uma das maiores virtudes da CART é a capacidade de pesquisa de relações entre os dados, mesmo que não sejam evidentes.

O método CART baseia-se na execução de partições binárias sucessivas de uma amostra, com base nos resultados amostrados das variáveis independentes, buscando a constituição de subamostras internamente homogêneas. A classificação dessas subamostras é realizada conforme alguma medida descritiva e a predição de novos elementos, executada por meio da estrutura de classificação constituída (TACONELLI, ZOCCHI e DIAS, 2009).

Os componentes elementares do modelo são os nós e as regras de divisão (*splitting rules*). O primeiro nó de uma árvore é chamado de raiz e representa todo o conjunto de dados. Os nós terminais recebem o nome de folhas. Os nós que dão origem a outros são chamados de pais, enquanto aqueles gerados são denominados de filhos.

As ANN são uma técnica de processamento de informação inspirada pelo sistema nervoso humano. O cérebro humano processa informações de maneira diferente quando comparado a um computador convencional. Segundo Hakin (2001), o cérebro pode ser considerado um sistema de processamento de informação extremamente complexo, não linear e paralelo, com capacidade de organizar seus componentes estruturais, os neurônios, para realizar atividades como reconhecimento de padrões, percepção e controle motor, executando-as de maneira muito mais eficaz que sistemas computacionais.

O primeiro modelo de ANN foi proposto por McCulloch e Pitts (1943), que propuseram um sistema para reproduzir as características básicas de um neurônio. O modelo McCulloch-Pitts é formado por uma série de entradas $E_1, E_2, \dots, E_n$ que dão origem a um valor binário y :

$$S = \sum_{k=1}^n E_k P_k$$

$$y = f(S)$$

em que $f(S)$ denota a função de ativação de Heaviside, assumindo o valor 1, caso S seja maior ou igual a zero, e valor nulo, caso S seja menor que zero. P_k representa os pesos associados às sinapses. Em caso de peso positivo, as sinapses são denominadas de excitação, caso o peso seja negativo, a sinapse é de inibição.

Segundo Corrar *et al* (2012), as ANN têm sido muito aplicadas na área de negócios, com objetivos cada vez mais diversificados, e encontrando, muitas vezes, resultados superiores aos métodos estatísticos convencionalmente aplicados. A técnica tem complementado e enriquecido modelos estatísticos de inadimplência, riscos securitários e avaliação dos riscos associado aos papéis financeiros, entre outras aplicações.

Na quarta etapa da metodologia proposta tem-se a comparação da habilidade preditiva de cada técnica de classificação empregada. Para isso, faz-se uso das curvas ROC, do inglês *receiver operating characteristic curves*. A escolha desse método em detrimento das precisões gerais obtidas justifica-se pois as classes que compõem a amostra não estão equilibradas, isto é, há diferenças entre a quantidade de empresas em cada um dos grupos (21 insolventes e 66 solventes).

Esse tipo de análise baseia-se na relação entre a sensibilidade e a especificidade de classificadores binários. A sensibilidade é a taxa de verdadeiros positivos (*true positive rate*), que nesse estudo é a probabilidade de uma empresa solvente ser classificada como tal. O segundo atributo, a especificidade, é a taxa de verdadeiros negativos (*true negative rate*), a probabilidade de uma empresa insolvente ser classificada nesse grupo.

De acordo com Flach (2010), se um modelo de classificação estima um escore que seja proporcional ao grau de certeza com o qual determinada entrada pertença à classe positiva, neste caso, de ser solvente, é possível determinar vários pontos de corte, os quais definirão diferentes proporções para os valores de sensibilidade e especificidade dos modelos. Observando todos os possíveis pontos de corte desde zero até um e ligando-os todos, forma-se uma “curva” composta por segmentos de reta, a qual recebe o nome de curva ROC.

A principal medida de desempenho que se pode obter desse gráfico é o que se chama de AUC, do inglês *area under the curve*, que consiste em calcular o valor da área abaixo da curva. A AUC pode ser interpretada como a probabilidade de uma observação positiva retirada aleatoriamente receba uma escore maior do que uma observação negativa retirada da mesma maneira. Quanto melhor o ajustamento do modelo ao conjunto de dados, maior será a AUC, que varia entre zero e um. Esta foi a medida adotada neste trabalho para avaliar o desempenho das técnicas utilizadas para a classificação dos dados.

Engelmann, Hayden e Tasch (2003) propõem uma transformação simples ao valor obtido para a AUC, cujo resultado denominam de taxa de precisão, do inglês *accuracy ratio* (AR). Para a obtenção dessa medida basta subtrair 0,5 do valor obtido pra a AUC, e então multiplicar o que se obtém por dois. Segundo os autores demonstram em seu artigo, essa transformação linear resulta na *Cumulative accuracy profile*, outra método de avaliação da capacidade preditiva de modelos de classificação bastante utilizado. Esta métrica também será considerada para a comparação do desempenho das técnicas estatísticas empregadas nesse estudo.

Também será empregado o teste de *Wilcoxon-Mann-Whitney* para que se avalie as diferenças entre os grupos de empresas solventes e insolventes. Esse teste será utilizando tanto para os indicadores contábeis, bem como para os escores preditos pelos modelos, sendo esta também uma forma de avaliar a qualidade das técnicas empregadas, conforme visto em Jackson e Wood (2013).

Os resultados obtidos para as técnicas de classificação foram validados utilizando o método de validação cruzada conhecido como *leave one out*, que consiste na extração de uma observação do conjunto de dados, seguida da aplicação da técnica estatística a ser utilizada, sendo o modelo proposto testado na observação que foi retirada inicialmente. O processo é repetido até que toda a amostra tenha sido utilizada na etapa de validação. Esse é um método indicado para amostras pequenas, como é o caso deste estudo.

As informações contábeis usadas nesse trabalho foram coletadas em 21 e agosto de 2013, por meio da ferramenta *Economatica*. Os dados foram tratados, inicialmente, em uma

planilha do Microsoft Excel 2010. Todas as análises estatísticas foram feitas com uso de pacotes do *software R: A language and enviroment for statistical computing* (R CORE TEAM, 2013).

4 Análise dos Resultados

4.1 Análise Descritiva e Fatorial

Seguindo a metodologia proposta, a primeira etapa da análise buscou extrair informações da base de dados a partir de estatísticas descritivas. A Tabela 1 traz a distribuição de frequência das empresas agrupadas segundo seus setores considerados.

Tabela 1 – Distribuição das empresas por setor.

Subsetores	Insolventes		Solventes		Total	
Agroindústria	1	4,76%	4	6,06%	5	5,75%
Comércio de máquinas e veículos pesados	1	4,76%	2	3,03%	3	3,45%
Construção e engenharia	2	9,52%	7	10,61%	9	10,34%
Eletroeletrônicos	1	4,76%	5	7,58%	6	6,90%
Energia elétrica	2	9,52%	10	15,15%	12	13,79%
Madeira e papel	1	4,76%	4	6,06%	5	5,75%
Material de construção	1	4,76%	1	1,52%	2	2,30%
Material de transporte	1	4,76%	5	7,58%	6	6,90%
Metalurgia	1	4,76%	4	6,06%	5	5,75%
Química	1	4,76%	4	6,06%	5	5,75%
Tecidos, vestuário e calçados	8	38,10%	18	27,27%	26	29,89%
Transportes aéreos	1	4,76%	2	3,03%	3	3,45%
Total	21	100,00%	66	100,00%	87	100,00%

Fonte: Elaborada pelo autor.

Percebe-se uma concentração de empresas do setor de tecidos, vestuário e calçados. Essa situação pode ser justificada por diversos fatores, como a crise do algodão, que começou a ser sentida no Brasil a partir de 2006. A estrutura familiar de boa parte das empresas que compõem esse segmento da economia, em especial as indústrias têxteis. E a forte concorrência chinesa.

A Tabela 2 apresenta medidas de tendência central, variabilidade e de posicionamento para os ativos totais dos dois grupos.

Tabela 2 – Medidas do tamanho do Ativo Total por grupo.

	Insolventes	Solventes	Geral
Mínimo	23.620	68.216	23.620
Máximo	12.935.830	13.662.280	13.662.280
Média	1.227.364	2.053.753	1.854.280
Mediana	182.075	574.097	506.987
Desvio	2.915.823	3.207.672	3.143.312

Fonte: Dados da pesquisa.

Como havia na amostra empresas insolventes no mesmo segmento, o Ativo Total foi utilizado como segundo critério de emparelhamento. Com a Tabela 2, percebe-se que, em geral, o grupo de empresas solventes apresenta maior Ativo Total, bem como maior variabilidade, considerando o desvio padrão.

Já na Tabela 3, são apresentadas estatísticas descritivas dos indicadores contábeis. A tabela traz medidas de variabilidade, tendência central e posicionamento. É apresentada ainda os resultados do teste (valor p) de Wilcoxon-Mann-Whitney, utilizado para testar a diferença entre os grupos de empresas solventes e insolventes para cada indicador.

Tabela 3 - Estatísticas descritivas das variáveis e teste de Wilcoxon-Mann-Whitney

Var.	Insolventes					Solventes					W
	1º Quart.	Med.	Média	3º Quart.	Desv.	1º Quart.	Med.	Média	3º Quart.	Desv.	
X1	-0,983	-0,391	-1,082	-0,170	1,966	0,038	0,225	0,160	0,317	0,247	0,000
X2	-0,212	-0,086	-0,102	0,070	0,196	0,037	0,082	0,094	0,124	0,108	0,000
X3	-0,631	-0,320	-0,185	0,303	0,516	0,554	0,800	1,040	1,316	0,888	0,000
X4	0,465	0,589	0,700	0,744	0,541	0,519	0,777	0,868	1,140	0,462	0,058
X5	0,767	1,470	2,216	2,712	2,218	0,422	0,535	0,563	0,619	0,271	0,000
X6	0,160	0,288	0,468	0,560	0,472	1,134	1,718	1,783	2,177	0,916	0,000
X7	-0,506	-0,194	-0,345	-0,053	0,369	0,010	0,042	0,041	0,076	0,128	0,000
X8	0,114	0,180	0,360	0,486	0,424	0,980	1,190	1,321	1,543	0,721	0,000
X9	0,152	0,310	0,453	0,608	0,406	0,746	1,178	1,208	1,457	0,641	0,000
X10	-0,950	-0,415	-0,650	0,182	1,662	0,605	1,466	2,072	3,156	2,194	0,000
X11	-0,735	-0,477	-1,346	-0,088	3,130	0,025	0,062	0,047	0,100	0,176	0,000
X12	-0,107	-0,042	-0,055	0,040	0,168	0,060	0,174	0,189	0,282	0,197	0,000
X13	0,001	0,005	0,029	0,041	0,042	0,056	0,209	0,366	0,387	0,537	0,000
X14	0,015	0,059	0,094	0,131	0,109	0,037	0,127	0,123	0,182	0,098	0,182
X15	0,004	0,012	0,059	0,094	0,101	0,036	0,140	0,474	0,402	1,391	0,000
X16	-0,343	-0,107	0,899	1,682	2,530	0,570	0,795	1,050	1,349	1,289	0,018

Fonte: Dados da pesquisa.

Ao se observar a tabela, nota-se algumas peculiaridades. No grupo das empresas solventes, todos os indicadores apresentam sinal positivo, enquanto nas insolventes há alguns com sinais negativos, valendo o mesmo raciocínio para a mediana e as demais medidas de posicionamento, é um indício da diferença entre os dois grupos. Corroborando com esse raciocínio, os resultados do teste de Wilcoxon mostram que das 16 variáveis, apenas três não apresentam diferenças estatisticamente significativas considerando um nível de 1%. São elas: Receita Líquida/Ativo Total (X_4), Estoques/Ativo Total (X_{14}) e Ativo Permanente/Patrimônio Líquido (X_{16}). Assim, há um indício de que os indicadores selecionados possuem de fato poder discriminatório para empresas solventes e insolventes.

Os valores obtidos para o teste de esfericidade de Bartlett e do KMO indicam que é possível aplicar a AF ao conjunto de dados, já que o primeiro apresentou um valor p inferior a 0,0001, enquanto o KMO obtido foi de 0,6201, considerado mediano segundo os padrões definidos por Kaiser (1974). Dessa forma, prosseguiu-se com a análise. O critério da raiz latente levou à escolha de oito fatores, dos quais retirou-se a variável com maior escore.

O critério da raiz latente levou à escolha de oito fatores, dos quais retirou-se a variável com maior escore. O total da variância explicada pelos oito fatores é de 92,10%. Todos os valores obtidos para as comunicações foram superiores a 0,5 (o menor valor observado foi de 0,7492 para a variável X_4). Já pelo critério da variável substituta, aquelas que foram escolhidas para serem utilizadas nas técnicas de classificação foram X_1 , X_3 , X_{11} , X_{12} , X_{13} , X_{14} , X_{15} e X_{16} .

4.3 Resultado das Técnicas Aplicadas

As precisões obtidas pelas técnicas empregadas no estudo, considerando um ponto de corte ótimo, são apresentadas na Tabela 4. Visando a atender o objetivo específico de ver quais variáveis são consideradas mais importantes para o estudo, foram observados as equações obtidas para os modelos de LDA e RL, bem como a árvore do modelo CART. Observou-se que o Capital Circulante Líquido sobre o Ativo Total (X_1) foi a única variável considerada importante pelas três técnicas simultaneamente. Essa variável é constantemente

reportada como uma das mais importantes para a previsão de insolvência, conforme pode ser visto em Beaver (1966), Altman (1968), Altman, Baydía e Dias (1979), dentre outros. Também foram consideradas importantes o Endividamento Geral (X_3) e Margem Líquida (X_{11}).

Tabela 4 - Precisões obtidas pelos modelos

Técnica	Precisão Tipo I	Precisão Tipo II	Precisão Geral
LDA	71,43%	98,48%	91,95%
RL	85,71%	92,42%	90,80%
kNN	71,43%	92,42%	87,36%
CART	80,95%	98,48%	94,25%
ANN	100,00%	98,48%	98,85%

Fonte: Dados da pesquisa.

Com base nas precisões obtidas é possível notar que o modelo de ANN obteve o melhor desempenho nos três tipos de precisão, o que corrobora com os resultados obtidos por estudos realizados anteriormente, que indicam que as ANN tem desempenho mais robusto, especialmente quando a amostra utilizada é pequena, como é o caso desta pesquisa. O método errou apenas a classificação de uma empresa no grupo das solventes, exatamente a empresa 59, que, como visto anteriormente, passou por graves dificuldades financeiras no período analisado. O modelo que obteve o segundo melhor desempenho pelo critério adotado foi o das CART, que errou no grupo de empresas solventes apenas a classificação da empresa 59, porém classificou erradamente quatro empresas no grupo das insolventes.

O modelo com pior desempenho foi, sem dúvida, o kNN, que, além de obter a precisão geral mais baixa, também obteve a pior precisão dentro de cada grupo. Não fica muito claro qual o modelo obteve melhor desempenho entre LDA e RL. O primeiro modelo, apesar de ter obtido um erro geral menor, especialmente se for levado em consideração que o único erro apresentado para o grupo de empresas solventes foi de uma empresa com problemas financeiros, sua precisão pode ter sido afetada pelo desbalanceamento dos grupos, uma vez que seus erros concentram-se nas empresas insolventes, que compõem a menor parte da amostra. Neste sentido, a RL apresenta uma distribuição de erro por grupos menos desproporcional.

Após a análise da precisão, procedeu-se então com a criação das curvas ROC. Para isso, foi feita inicialmente a coleta das probabilidades *a posteriori*, ou seja, das probabilidades preditas de que cada observação pertencesse ao grupo de empresas solventes (classe positiva). Ressalta-se que, como o modelo de kNN considerou apenas um vizinho, as probabilidades dadas pelo método de uma observação ser solvente só poderiam ser 0 ou 1, isto é, o vetor de probabilidades é igual ao vetor de classificação. A Tabela 5 mostra os resultados extraídos com base na curva ROC, bem como o resultado do teste de Wilcoxon (valor *p*) para as diferenças entre os escores de empresas solventes e insolventes calculados por cada técnica.

Tabela 5 – Resultados para a AUC, AR e Teste de Wilcoxon

Técnica	AUC	AR	W
LDA	0,9221	0,8442	0
RL	0,9199	0,8398	0
kNN	0,8193	0,6386	0
CART	0,9369	0,8738	0
ANN	0,9895	0,9790	0

Fonte: Dados da pesquisa.

O melhor modelo nesse quesito foi novamente o construído com as ANN, cuja área abaixo da curva (AUC) foi de 0,9906. As CART novamente obtiveram o segundo melhor desempenho, com AUC de 0,9369. O desempenho da LDA e da RL foi bastante parecido,

com uma ligeira vantagem para o primeiro modelo, cuja área calculada foi de 0,9221, enquanto o segundo obteve 0,9199. O único modelo com AUC inferior à 0,9 foi o kNN, cuja área calculada foi de 0,8193. Segundo os critérios normalmente adotados para a análise da AUC, valores acima de 0,9 são tidos como excelentes, enquanto valores entre 0,8 e 0,9 são considerados bons. Os resultados obtidos para o AR apenas corroboram com aqueles demonstrados para a AUC, cabendo ressaltar a superioridade das redes neurais (ANN), que fica mais evidente ao se analisar as estatísticas AR. Os resultados do teste de Wilcoxon também confirmam a qualidade das técnicas empregadas, uma vez que se evidencia que há diferença estatisticamente significativa entre os grupos no que diz respeito à probabilidade de vir a se tornar insolvente. Os resultados obtidos se assemelham àqueles observados por Jackson e Wood (2013) no tocante aos valores obtidos tanto para a AUC, como para a AR, e também na técnica estatística considerada superior, uma vez que os autores também encontraram que as ANN, dentre os modelos que utilizam dados contábeis, como superiores aos demais.

5 Conclusão

O presente trabalho apresentou como objetivo principal avaliar o desempenho de técnicas estatísticas de classificação aplicadas à problemática da previsão de insolvência de empresas. Para tanto foram construídos modelos de classificação com base em cinco técnicas de classificação: análise discriminante, regressão logística, classificador dos vizinhos mais próximos, árvores de classificação e regressão e redes neurais artificiais. Os resultados obtidos para cada um dos modelos foram submetidos a uma análise de curvas ROC, com o intuito de se extrair informações sobre seus desempenhos.

Com base nos resultados é possível concluir que modelos de classificação constituem uma ferramenta poderosa para prever problemas financeiros, podendo auxiliar gestores e investidores na tomada de decisão e contribuindo para a redução do risco de crédito. O modelo baseado em Redes Neurais obteve um desempenho consideravelmente superior aos demais, sendo este resultado esperado com base no que se apresenta na literatura.

Este trabalho apresentou como limitação, o tamanho reduzido da amostra, que contava com apenas 87 empresas. Outro ponto a ser destacado foi a ausência de algumas informações nas demonstrações contábeis o que dificultou ou impossibilitou a construção de alguns indicadores. Outra dificuldade encontrada foram as mudanças ocorridas nas normas contábeis ao longo do período estudado, que implicaram, não raro, em mudanças na estrutura das demonstrações contábeis.

Existem ainda limitações ligadas à utilização de indicadores contábeis para esse tipo de estudo. A primeira delas é a possibilidade de manipulação por parte dos gestores, o que pode gerar índices que não representam de forma fidedigna a realidade da empresa. Há também o fato de que as demonstrações contábeis de um determinado ano apenas serão divulgadas durante o ano subsequente. Como neste trabalho foram utilizadas demonstrações de um ano anterior à entrada em insolvência, algumas das empresas já haviam feito seu pedido de recuperação judicial antes de divulgarem as informações contábeis necessárias para a elaboração dos índices, fato que ocorreu em quatro das 21 empresas insolventes utilizadas no estudo.

Para estudos futuros, sugere-se a aplicação de métodos de classificação diferentes, como a máquina de suporte vetorial, a análise por envoltória de dados e o *random forest*. Também podem ser utilizadas diferentes técnicas para a seleção das variáveis como as abordagens de filtro e *wrapper*, ao invés da análise fatorial, que por sua vez, também pode ser utilizada de forma diferente, sendo considerado os escores obtidos pelas empresas em cada dimensão, e não o valor da variável substituta. Sugere-se também a adoção de outras

perspectivas, como o efeito da adoção do IFRS sobre a capacidade preditiva das técnicas de previsão de insolvência.

Referências

- ALTMAN, Edward I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. **Journal of Finance**, Boston, v.23, p. 586 - 609. 1968.
- ALTMAN, E. I.; BAIDYA, T. K. N.; DIAS, L. M. R. Previsão de problemas financeiros em empresas. **Revista de Administração de Empresas**, Rio de Janeiro, v. 19, n. 1, p. 17-28, 1979.
- AITA, Jaqueline; ZANI, João; SILVA, Carlos E. S. Determinantes de insolvência bancária no Brasil: identificação de evidências macro e microeconômicas. *In*: XI Encontro Brasileiro de Finanças, 2011, Rio de Janeiro. **Anais...** Rio de Janeiro: FGV, 2011.
- BASGALUPP, Márcio Porto. **LEGAL-Tree: Um algoritmo genético multi-objetivo lexicográfico para indução de árvores de decisão**. Tese (Doutorado em Ciências da Computação e Matemática Computacional) – Universidade de São Paulo, São Paulo, 2010.
- BEAVER, William H. Financial Ratios as Predictors of Failure. **Journal of Accounting Research**, Chicago, Supplement of Selected Studies, p. 77-111. 1966.
- BRITO, Giovani A. S.; ASSAF NETO, Alexandre. Modelo de classificação de risco de crédito de empresas. **Revista de Contabilidade e Finanças**, São Paulo, v. 19, n. 46, p.18-29, 2008.
- BRITO, Giovani A. S.; ASSAF NETO, A.; CORRAR, L. J. Sistema de classificação de risco de crédito: uma aplicação a companhias abertas no Brasil. **Revista de Contabilidade e Finanças**, São Paulo, v. 20, n. 51, p.28-43, 2009.
- CARVALHO, Flávio Leonel et al. Identificação de indicadores contábeis relevantes para previsão e projeção de rentabilidade. **Revista de Educação e Pesquisa em Contabilidade**, Brasília, v. 4, n. 3, p. 94-110, 2010.
- CASTRO JÚNIOR, F. H. F. **Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais**. Dissertação (Mestrado em Administração) – Universidade de São Paulo, São Paulo, 2003.
- CHUNG, Kim-Choy; TAN, S. S.; HOLDSWORTH, D. K. Insolvency prediction model using multivariate discriminant analysis and artificial neural network for finance industry in New Zeland. **International Journal of Business and Management**, Toronto, v. 3, n. 1, 2008.
- CLARO, Carlos Roberto. **Recuperação Judicial: Sustentabilidade e função social da empresa**. Dissertação (Mestrado em Direito Empresarial e Cidadania), Centro Universitário de Curitiba, Curitiba, 2008.
- CORRAR, Luiz J.; PAULO, Edilson; DIAS FILHO, José Maria. **Análise Multivariada para os cursos de Administração, Ciências Contábeis e Economia**. São Paulo: Atlas, 2012.
- ELISABETSKY, Roberto. **Um modelo matemático para decisões de crédito no Banco Comercial**. 1976. Dissertação (Mestrado) – Instituto Politécnico da Universidade de São Paulo, 1976.
- ENGELMANN, Bernd; HAYDEN, Evelyn; TASCH, Dirk. Testing rating accuracy. **Risk**, jan., 2003.
- FITZPATRICK, Paul J. **A Comparison of the Ratios of the Successful Industrial Enterprises with those of Failed Companies**. The Accountants Publishing Company, 1932.

FLACH, Peter A. ROC Analysis. **Encyclopedia of Machine Learning**. Nova Iorque: Springer, 2010, 1031 p.

GUIMARÃES, Ailton; MOREIRA, Tito B. S. Previsão de insolvência: um modelo baseado em índices contábeis com utilização da análise discriminante. **Revista de Economia Contemporânea**, Rio de Janeiro. v. 12, n. 1, p. 151-178, jan./abr. 2008.

HAIR JUNIOR, J. F. et al. **Análise multivariada de dados**. Porto Alegre: Bookman, 2005.

HAYKIN, Simon. **Redes Neurais: Princípios e prática**. Porto Alegre: Bookman, 2001.

HORTA, Rui A. Mathiasi. **Uma metodologia de mineração de dados para a previsão de insolvência de empresas brasileiras de capital aberto**. Tese (Doutorado em Engenharia Civil) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

HORTA, Rui. A. Mathiasi. *et al.* Comparação de técnicas de seleção de atributos para previsão de insolvência de empresas brasileiras no período 2005-2007. *In: Encontro Nacional de Pós-Graduação e Pesquisa em Administração*, 34, 2010, Rio de Janeiro. **Anais...** Rio de Janeiro: Anpad, 2010.

JACKSON, R. H. G.; WOOD, A. The performance of insolvency prediction and credit risk models in the UK: a comparative study. **The British Accounting Review**, v. 45, p. 183-202, 2013.

KAISER, Henry F. An index of factorial simplicity. **Psychometrika**, v. 39, n. 1, p. 31-36, 1974

KANITZ, Stephen Charles. **Como prever falências**. São Paulo, Mc Graw-Hill, 1978.

LACHTERMACHER, Gerson; ESPENCHITT, Dilson G. Previsão de falência de empresas: estudo de generalização de redes neurais. *In: Encontro Nacional de Pós-Graduação e Pesquisa em Administração*, 25, 2001, Campinas. **Anais...** São Paulo: Anpad, 2001.

MATIAS, Alberto B. **Indicadores contábeis e financeiros de previsão de insolvência: a experiência da pequena e média empresa**. Tese (Livre-Docência) – Universidade de São Paulo, São Paulo, 1976.

MINUSSI, João A.; DAMASCENA, Cláudio; NESS JR, Walter L. Um modelo de previsão de solvência utilizando regressão logística. **Revista de Administração Contemporânea**, Curitiba. v. 6 n. 3 p. 109-128, set./dez. 2002.

OHLSON, James A. Financial Ratios and the Probabilistic Prediction of Bankruptcy. **Journal of Accounting Research**, Chicago, v. 18, n. 1, 1980.

ONUSIC, Luciana Massaro, et al; Estudo exploratório utilizando as técnicas de análise por envoltória de dados e redes neurais artificiais na previsão de insolvência de empresas. **FACEF Pesquisa**, v. 9, n. 2, 2006.

PEDRO, Sílvia M. D. **Exploração de dados aplicada à análise de risco de crédito**. Monografia (Licenciatura em Matemática Aplicada e Computação) – Instituto Superior Técnico, Lisboa, 2001.

REBOUÇAS, S. M. D. P. **Metodologias de classificação supervisionada para análise de dados de microarrays**. Tese (Doutorado em Estatística e Investigação Operacional) – Universidade de Lisboa, Lisboa, 2011.

SANVICENTE, Antônio Z.; MINARDI, Andrea M. A. F.; Identificação de indicadores contábeis significativos para previsão de concordata de empresas. **Finance Lab Working Papers**, IBMEC, São Paulo, 1998.

SCARPEL, Rodrigo A. Previsão de insolvência de empresas utilizando support vector machine. **Revista de Economia e Administração**, São Paulo, v. 7, n. 3, p. 281-295, 2008.

SILVA, Rômulo O. R. C. **Estudo de insolvência de empresas de capital aberto**. Trabalho acadêmico (Mestrado em Administração) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2006.

SILVA, José Pereira da. **Modelos para classificação de empresas com vistas à concessão de crédito**. Dissertação (Mestrado em Administração) – Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo, 1982.