



AValiação da Acurácia de Classificadores Aplicada à Predição de Insolvência de Companhias Não Financeiras Listadas na BM&FBOVESPA

ARTIGO ORIGINAL

NETO, Fernando Henrique Rodrigues ¹

PESSANHA, José Francisco Moreira ²

DUQUE, Andréa Paula Osório ³

ALVES, Francisco José dos Santos ⁴

NETO, Fernando Henrique Rodrigues. Et al. **Avaliação da acurácia de classificadores aplicada à predição de insolvência de companhias não financeiras listadas na BM&FBOVESPA**. Revista Científica Multidisciplinar Núcleo do Conhecimento. Ano 04, Ed. 08, Vol. 06, pp. 72-100. Agosto de 2019. ISSN: 2448-0959, Link de acesso: <https://www.nucleodoconhecimento.com.br/contabilidade/avaliacao-da-acuracia>

RESUMO

Este trabalho objetiva comparar as acurácias obtidas para cada um dos seguintes classificadores: análise discriminante (LDA), regressão logística (RL), redes neurais – *multilayer perceptron* (MLP) e máquinas de vetores de suporte (SVM), em diferentes períodos. Com a globalização da economia, o aumento da disponibilidade de

¹ Mestre em Ciências Contábeis pela UERJ, Especialista em Gestão Financeira pela UCAM, Graduado em Ciências Contábeis e em Administração pelo Ibmecc-RJ.

² Doutor em Engenharia Elétrica (PUC-Rio).

³ Doutora em Ciência da Informação (UFF).

⁴ Doutor em Ciências Contábeis (USP).



informações financeiras e o livre fluxo de capitais ao redor do mundo, investidores, profissionais de mercado, acadêmicos, dentre outros usuários da informação contábil, vêm se preocupando em saber se empresas estão com dificuldades financeiras. A pesquisa realizada é descritiva e parte de uma abordagem quantitativa. As variáveis independentes utilizadas correspondem a um conjunto de indicadores financeiros calculados com base em dados oriundos de demonstrações financeiras de companhias listadas na BM&FBOVESPA. A seleção das variáveis foi feita por meio do método *stepwise* e algoritmo genético. Nos estudos realizados, a regressão logística apresentou o melhor desempenho nos grupos de teste de 2015 e 2016. Essa pesquisa contribuiu para preencher um *gap* em relação à utilização de dois grupos de treinamento para averiguar se os dados contábeis do ano de implementação plena das *IFRS* (2010) no Brasil podem ter impactado as acurácias dos classificadores.

Palavras-chave: Previsão de insolvência, indicadores financeiros, classificadores.

INTRODUÇÃO

Com a evolução do mercado de capitais e de crédito, além do maior fluxo de informações contábeis disponíveis, investidores necessitam saber se uma companhia corre risco de insolvência. Dessa forma, cada vez mais, investidores e profissionais de mercado estão dando atenção ao tema da previsão de insolvência de empresas, em nível internacional (Altman; Baidya; DIAS, 1979). Tal assunto se torna relevante, pois os interessados em investir em instituições com fins lucrativos podem ter o intuito de saber se determinada entidade corre o risco de encerrar suas atividades operacionais. Para um potencial acionista, é importante saber se uma companhia possui capacidade de continuidade para realizar um determinado investimento. Já para um credor, pode ser pertinente investigar se uma organização é capaz de honrar seus compromissos ou se está sob risco de insolvência.

Alguns profissionais, como analistas e gestores de crédito, acreditam que pode ser interessante a utilização de modelos de previsão de insolvência. Inclusive, no processo de avaliação de crédito, utilizam-se técnicas para selecionar quem deve recebê-lo (Gitman, 2007). Atualmente, pesquisadores vêm utilizando classificadores



como meio para prever se as organizações estão em estado de insolvência. Todavia, há uma variedade de metodologias disponíveis para esse fim. Considerando tais informações, chegou-se ao seguinte problema de pesquisa: qual é o método mais acurado para prever insolvência de companhias não financeiras listadas na BM&FBOVESPA: análise discriminante, regressão logística, redes neurais ou máquinas de vetores de suporte? Sendo assim, o objetivo deste estudo consiste em comparar as acurácias obtidas para cada um dos classificadores com base em indicadores financeiros.

A partir dos resultados, os interessados em saber sobre o potencial de uma empresa vir a se tornar insolvente poderá utilizar um classificador em detrimento de outro, em suas análises. Este artigo se diferencia dos demais por apresentar um recorte temporal (2010 a 2016) que se dispõe a discorrer sobre as mudanças nas normas contábeis no Brasil devido à implementação das normas internacionais, principalmente a partir de 2010 (Pasini, 2015). Sendo assim, pretende-se averiguar se as acurácias dos classificadores apresentarão similaridade nos anos de 2015 e 2016, sendo estes os períodos de teste. Somado a isso, os grupos de treinamento serão divididos em duas partes e em períodos distintos. O primeiro contemplará o intervalo de 2010 a 2014, enquanto o segundo, de 2011 a 2014, uma vez que 2010 foi o ano de implementação plena das IFRS no Brasil (SANTOS; PONTE; Mapurunga, 2014). Dessa forma, este trabalho busca preencher esse *gap* existente no que tange a predição de insolvência de empresas.

1. REFERENCIAL TEÓRICO

1.1 INSOLVÊNCIA

De forma genérica, insolvência é o estado em que uma empresa possui um passivo maior que o ativo, ou seja, tem menos bens e direitos que o total de dívidas (Coelho, 2010). Corroborando tal conceito, Ross, Westerfield e Jaffe (2013) salientam que o estado de insolvência de uma entidade ocorre "quando uma empresa possui patrimônio líquido negativo, de modo que o valor dos ativos é inferior ao valor das dívidas". Isso quer dizer, conforme asseveram Ross *et al.* (2013), que a insolvência



pode levar a um problema de incapacidade de honrar dívidas, já que se todos os passivos forem exigidos, uma entidade não teria ativos suficientes para poder sanar seus compromissos financeiros.

De forma análoga, uma empresa pode ser considerada como insolvente caso tenha iniciado um processo de falência em um dado espaço de tempo (Brito; Assaf, 2008). Contudo, as companhias podem estar em estado de insolvência, ou seja, com passivos exigíveis maiores que os ativos e não terem dado entrada em processo de falência. Sendo assim, é importante que os investidores saibam quais são as empresas que possuem perfil de insolvência. É um processo necessário e indispensável, pois, dessa forma, eles podem tomar suas decisões de maneira mais adequada, mitigando, assim, o risco de tal infortúnio.

1.2 INDICADORES FINANCEIROS

Os trabalhos pioneiros sobre o estudo de infortúnios empresariais com base em indicadores financeiros, obtidos por meio de dados contábeis de demonstrações financeiras, começaram na década de 1930 com Fitzpatrick (1932). Contudo, foi no fim da década de 1960 e durante a de 1970 que mais estudos foram difundidos sobre o tema, utilizando, dessa forma, novas técnicas. Alguns autores que fizeram tais contribuições nesse período foram: Beaver (1968), Altman (1968), Deakin (1972), Edmister (1972), Elizabetsky (1976), Kanitz (1978), Matias (1978), dentre outros. Apesar do desenvolvimento de novas metodologias para a área, os indicadores financeiros não foram descartados das análises mais modernas.

Tais indicadores têm o objetivo de evidenciar a situação econômica ou financeira de uma entidade por meio de relações entre contas das demonstrações financeiras. Contudo, não é possível obter um diagnóstico preciso pelos resultados que são gerados por tais índices, pois há outros fatores externos ao ambiente da empresa que influenciam seu desempenho (Matarazzo, 2010). Esses quocientes são utilizados, neste estudo, com o intuito de auxiliar na classificação das empresas estudadas, segregando-as em solventes e insolventes. Iudícibus (2009) salienta que a análise de



quocientes não é utilizada, apenas, para retratar o passado de uma entidade, mas também, para auxiliar inferências futuras.

Para isso, será realizada uma seleção de indicadores com maior representatividade para a investigação em cada análise. Para Matarazzo (2010), os indicadores podem ser de três tipos: liquidez, estrutura de capital e rentabilidade. Os dois primeiros são de natureza financeira e o último, econômica. Os índices de liquidez “evidenciam a situação financeira de uma empresa frente a seus diversos compromissos financeiros” (Assaf, 2012, p. 176). Já os de estrutura de capital “mostram as grandes linhas de decisões financeiras, em termos de obtenção e aplicação de recursos” (Matarazzo, 2010, p. 87). Os de rentabilidade, conforme Matarazzo, (2010, p. 87) procuram mostrar “qual a rentabilidade dos capitais investidos, isto é, quanto renderam os investimentos e, portanto, qual o grau de êxito econômico da empresa”.

1.3 MÉTODOS DE CLASSIFICAÇÃO

Neste trabalho, os métodos de classificação são empregados com o propósito de distinguir as empresas solventes e insolventes com base em um conjunto de variáveis explicativas. Para realizar tal distinção, os seguintes classificadores foram considerados: análise discriminante, regressão logística, redes neurais e máquinas de vetores de suporte. A escolha dos quatro métodos foi realizada porque há diferentes trabalhos envolvendo tais classificadores, mas nenhum contemplou, simultaneamente, os propostos neste estudo (Pereira; Martins, 2015).

1.3.1 ANÁLISE DISCRIMINANTE (LDA)

Trata-se de uma abordagem estatística tradicionalmente empregada em tarefas de classificação e segundo Matarazzo (2010, p. 167), é “[...] uma poderosa técnica estatística capaz de dizer se determinado elemento pertence a uma população X ou Y”. Para Malhotra (2001, p. 482), “a análise discriminante é uma técnica de análise de dados em que a variável dependente é categórica e as variáveis prognosticadoras ou independentes têm natureza intervalar”. Aplicada a este estudo, a análise discriminante busca gerar uma função matemática com indicadores financeiros que



consigam segregar as empresas, alocando-as nos seus respectivos grupos das solventes e das insolventes, conforme Matarazzo (2010).

Em relação a tal função, conforme Matarazzo (2010, p. 167 - 168), "a variável dependente é comparada a um número fixo - chamado ponto crítico predeterminado pelo modelo. Se o valor da variável dependente ficar acima do ponto crítico, a empresa estará entre as solventes; se ficar abaixo, entre as insolventes". A formulação matemática desta análise é a que se segue abaixo. Por meio dessa fórmula matemática é possível alocar as companhias em seus respectivos grupos, dependendo do resultado da variável dependente.

$$Y = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \dots \quad (1)$$

Em que:

Y= variável dependente; reflete o total de pontos alcançado pela empresa

a_1, a_2, a_3, a_4 = pesos: indicam a importância relativa de cada índice

x_1, x_2, x_3, x_4 = variáveis independentes: são os índices de balanço

1.3.2 REGRESSÃO LOGÍSTICA (RL)

Trata-se de um modelo de regressão em que a variável dependente é binária (Fávero et al 2009). Essa técnica, segundo Fávero *et al.* (2009, p. 440), "destina-se a investigar o efeito das variáveis pelas quais os indivíduos, objetos ou sujeitos estão expostos sobre a probabilidade de ocorrência de determinado evento de interesse". Conforme o autor, "a função logística assume valores entre 0 e 1, para qualquer z entre $-\infty$ e $+\infty$ ". A representação matemática da fórmula segue abaixo:



$$f(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

Sendo z :

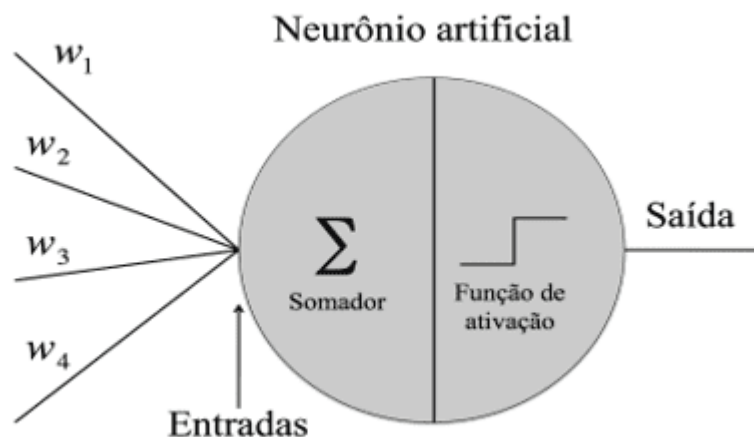
$$z = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_k X_k \quad (3)$$

O valor de z poderá mostrar se uma determinada entidade está situada no grupo de solventes ou insolventes. Desta forma, " p indica a probabilidade de ocorrência de determinado evento de interesse, X representa o vetor de variáveis explicativas (ou independentes) e α e β os parâmetros do modelo" (Fávero *et al.*, 2009, p. 441). Esses parâmetros são estimados por meio da análise de máxima verossimilhança.

1.3.3 REDES NEURAIS – MULTILAYER PERCEPTRON (MLP)

As redes neurais – *Multilayer Perceptron* (MLP) são modelos matemáticos inspirados no cérebro humano, pela sua complexidade, não linearidade e forma de processar os dados (Haykin, 1999). Este modelo utiliza entidades de processamento interconectadas que por meio dos estímulos (informações de entrada), é capaz de produzir uma resposta (predição). A unidade de processamento básica de uma rede neural é o neurônio, cuja estrutura e o funcionamento podem ser modeladas pelo neurônio artificial indicado na Figura 1.

Figura 1: Neurônio artificial de Mcculloch e Pitts



Fonte: O Autor (2018)

A equação que representa a saída de um neurônio individual é a seguinte:

$$y_i = f\left(\sum_j \beta_{ij} X + \alpha_{ij}\right) \quad (4)$$

Em que f é a função de ativação, os índices sob escritos i e j representam a interconexão entre o neurônio de entrada e o neurônio da camada escondida. β , α e X são a matriz de pesos, de termos independentes e o vetor de variáveis. A saída da rede é dada pelo mapeamento da soma da saída de cada um dos neurônios. Conforme o processo de aprendizagem ocorre, os pesos das interconexões vão sendo ajustados aos dados visando minimizar o erro entre a entrada e saída (Haykin, 1999).

O algoritmo de treinamento utilizado foi o *backpropagation*. Conforme Haykin, (1999), este consiste das seguintes etapas:

1. Inicializar os pesos (w) da rede aleatoriamente
2. Calcular o valor de cada uma das saídas referentes aos respectivos exemplos de treinamento usando a Equação 5.



$$\hat{y} = \sum_{i=0}^m w_i(t) \cdot y_i(t) \quad (5)$$

em que y_i é a saída de um neurônio específico da camada escondida dada pela Equação 4 e w_i é o peso entre as conexões do neurônio i com o neurônio da camada de saída.

3. Calcular os erros entre as saídas previstas pela Equação 5 e os valores exatos y_{exa} .

$$e = \sum \frac{1}{2} (\hat{y} - y_{exa})^2 \quad (6)$$

4. Retropropagar o erro estimado pela Equação 6, utilizando a Equação 7 se o índice l representar a camada de saída e Equação 8 caso o mesmo for referente à outra camada:

$$\delta_j^o = f'_j(s_j)(\hat{y} - y) \quad (7)$$

$$\delta_j^l = f'_j(s_j) \sum_{k \in out(j)} \delta_k^{l+1} w_{jk} \quad (8)$$

em que δ_j^l representa o erro da camada l , f' é a função de ativação e s_j o sinal de entrada no neurônio j .

5. Atualizar os pesos utilizando a Equação 9.

$$w_{jk}(t+1) = w_{jk}(t) - \eta \delta_j^{l+1}(t) y_j^l(t) \quad (9)$$



6. Avaliar os critérios de parada. Quando satisfeitos, finalizar processo de aprendizagem. Caso contrário, refazer os passos de 2 a 6 novamente até que os critérios sejam alcançados.

1.3.4 MÁQUINAS DE VETORES DE SUPORTE (SVM)

As máquinas de vetores de suporte (SVM) foram fundamentadas na teoria de aprendizagem de máquina (Hamel, 2009). Além disso, a metodologia é utilizada, também, para fazer tanto regressões quanto classificações binárias, como por exemplo, a avaliação de solvência das empresas, sendo esta, uma das proposições do presente trabalho. Uma das finalidades desse classificador é dividir os dados inseridos em dois grupos, por meio de um hiperplano que é utilizado para otimizar a classificação dos dados (Cristianini; Taylor, 2003). Schmueli, Patel e Bruce (2010) afirmam que o objetivo dessa divisão é buscar a maximização da distância entre os dados mais próximos para cada um dos grupos. Como salienta Lorena e Carvalho (2007), "classificadores que separam os dados por meio de um hiperplano são denominados lineares". Segundo Lorena e Carvalho (2007), A equação de um hiperplano é apresentada da seguinte forma:

$$f(X) = \beta \cdot X + \alpha = 0 \quad (10)$$

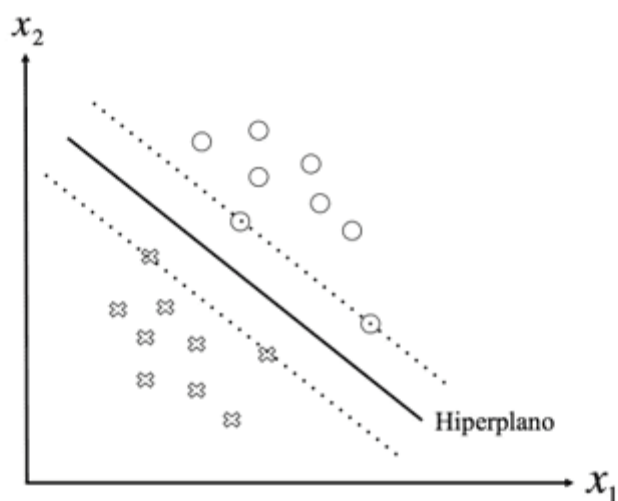
Sendo que, conforme Lorena e Carvalho (2007): " α e $\beta \cdot X$ é o produto escalar entre os vetores β e X , $\beta \in \Re$ é o vetor normal ao hiperplano descrito e $\alpha/||\beta||$ corresponde à distância do hiperplano em relação à origem, com $\alpha \in \Re$ ".

Além disso, a equação acima oferece a possibilidade de dividir o espaço dos dados coletados em duas partes. Sendo uma representada por $\beta \cdot X + \alpha > 0$ e a outra por $\beta \cdot X + \alpha < 0$. Sendo assim, como destaca Lorena e Carvalho (2007), "uma função sinal $g(X) = \text{sgn}(f(X))$ pode então ser empregada na obtenção das classificações". Abaixo segue a formulação de $g(X)$, de acordo com Lorena e Carvalho (2007):

$$g(X) = \text{sgn}(f(X)) = \begin{cases} 1 & \text{se } \beta \cdot X + \alpha > 0 \\ -1 & \text{se } \beta \cdot X + \alpha < 0 \end{cases} \quad (11)$$

A Figura 2 ilustra um hiperplano (reta) em duas dimensões que divide o espaço em duas classes, dadas pela Equação 11.

Figura 2: Hiperplano separador de classes



Fonte: O Autor (2018)

O treinamento de uma SVM baseia-se na solução do problema de otimização presente na Equação:

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|w\|^2 \\ \text{sujeito a} \quad & y_i(w \cdot \Phi(x_i) + b) \geq 1, \quad i = 1, 2, \dots, n, \end{aligned} \quad (12)$$

em que w é o vetor normal ao hiperplano, $\Phi(x_i)$ é a função de mapeamento que é aplicada aos dados de treinamento x_i e b é coeficiente linear do hiperplano definido no espaço de características. Este problema é tratado utilizando o método dos



multiplicadores de Lagrange, que modifica a função objetivo original por meio da adição ponderada das restrições, como mostra a Equação 13:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w} - \Phi(\mathbf{x}_i) + b) - 1) \quad (13)$$

A Equação 13 define a função Lagrangiana, que deve ser minimizada em relação às variáveis \mathbf{w} e b e maximizada em relação aos multiplicadores de Lagrange α_i . Dessa forma, tomam-se as derivadas parciais na Equação 13 e, em seguida, são igualadas a zero.

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \quad \text{e} \quad \frac{\partial L}{\partial b} = 0 \quad (14)$$

A solução da Equação 14 é dada pela Equação 15:

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad \text{e} \quad \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \Phi(\mathbf{x}_i) = 0 \quad (15)$$

Substituindo a Equação 15 na Equação 13, chega-se ao novo problema de otimização dado pela Equação 16:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) \\ \text{sujeito a} \quad & \alpha_i \geq 0, \quad i = 1, 2, \dots, n \\ & \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad (16)$$

em que $\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j)$ representa o produto escalar entre $\Phi(\mathbf{x}_i)$ e $\Phi(\mathbf{x}_j)$, denominada de função *kernel*. A solução do problema dual (Equação 16) é dada pela Equação 17:



$$\begin{aligned} f(\mathbf{x}) &= \text{sgn} \left(\sum_{i=1}^n \alpha_i y_i (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)) + b \right) \\ &= \text{sgn} \left(\sum_{i=1}^n \alpha_i y_i \mathbf{k}(\mathbf{x}, \mathbf{x}_i) + b \right) \end{aligned} \quad (17)$$

em que $f(\mathbf{x})$ é a função que define classificador ótimo treinado com os dados de interesse.

2. METODOLOGIA

O tipo da pesquisa é descritivo, pois busca definir qual é o classificador que possui melhor acurácia ao prever a situação de insolvência das companhias estudadas, conforme Cooper e Schindler (2003). A abordagem é quantitativa porque são utilizados métodos quantitativos com o intuito de classificar as companhias em solventes e insolventes. A população considerada neste trabalho é composta por companhias não financeiras listadas na BM&FBOVESPA, nos anos de 2010 a 2016. Esse intervalo de tempo foi escolhido porque em 2010 começou a adoção plena das normas IFRS no Brasil, tendo como exemplos a redução ao valor recuperável de ativos; ativo intangível; divulgação sobre partes relacionada; arrendamento mercantil financeiro para a arrendatária etc (Santos *et al.*, 2014).

Nos anos anteriores a 2010 era utilizado o BR GAAP e, por isso, não foi selecionado nenhum ano antecessor a 2010. Foi realizado um estudo por Santos *et al.* (2014) sobre a adoção e conformidade de divulgação das IFRS por parte das empresas não financeiras listadas na BM&FBOVESPA em 2010. Eles salientam que a aplicabilidade das normas internacionais foi parcial, já que o ano de 2010 foi um período de adaptação e aprendizagem em relação às mudanças. Portanto, por questões de precaução, foi utilizado o período de 2010 a 2016 e de 2011 a 2016, tendo em vista o cenário conturbado ocorrido em 2010. Na sequência, as empresas foram classificadas em dois grupos - solventes e insolventes.



A escolha dessas corporações foi devida aos dados serem públicos, além de auditados por empresas de auditoria que são credenciadas pela Comissão de Valores Mobiliários (CVM). No ano de 2010 foi obtido um total de 275 empresas, sendo 253 solventes e 22 insolventes. Já em 2011, havia 285 companhias, compostas por 261 solventes e 24 insolventes. Em 2012, o número de organizações era 295, com 267 solventes e 28 insolventes. No ano de 2013, o total era de 300 corporações, divididas em 272 solventes e 28 insolventes. O ano de 2014 apresentou 309 empresas, sendo 279 solventes e 30 insolventes.

No penúltimo ano de investigação, 2015, o total de companhias era de 307, separadas por 264 solventes e 43 insolventes. Já no último ano, 2016, havia 41 empresas insolventes e 260 solventes, totalizando 301 entidades. Cabe ressaltar que as companhias utilizadas na pesquisa estão dispostas no Anexo. Foram utilizados dados secundários retirados do banco de dados Economatica® que são aqueles que, de acordo com Malhotra (2001, p. 127), “já foram coletados para objetivos que não os do problema em pauta”. Tais dados da análise compõem as demonstrações financeiras, sendo que as utilizadas foram a Demonstração de Resultado do Exercício (DRE), Balanço Patrimonial (BP) e Demonstrações dos Fluxos de Caixa (DFC).

A partir dos dados secundários retirados das demonstrações financeiras das empresas analisadas, foram calculados os indicadores financeiros de interesse. Segundo o estudo de revisão feito por Pereira e Martins (2015), não há um consenso de um número exato de indicadores a serem utilizados para a predição de insolvência. Conforme sugere Pereira e Martins (2015), “uma área que parece ter pouca influência sobre as capacidades de previsão dos modelos é o número de indicadores considerados no modelo. Os modelos com eficiência de 100%, utilizaram entre 2 e 211 indicadores”. Decidiu-se, então, adotar o número médio de indicadores (doze) presentes nessa coletânea de trabalhos. Além disso, foram escolhidos aqueles que apresentaram maior frequência em estudos nacionais e internacionais, de acordo com Pereira e Martins (2015), conforme a Tabela 1.



Tabela 1: Indicadores e suas frequências em trabalhos nacionais e internacionais

Indicador	Frequência
Lucro Líquido	110
Total do Ativo	
Ativo Circulante	104
Passivo Circulante	
Capital de Giro	74
Total do Ativo	
Lucros Acumulados	72
Total do Ativo	
EBIT	54
Total do Ativo	
Vendas	54
Total do Ativo	
(Ativo Circulante-Estoques)	52
Passivo Circulante	
Dívida Total	41
Total do Ativo	
Ativo Circulante	40
Total do Ativo	
Lucro Líquido	39
Patrimônio Líquido	
Caixa	31
Total do Ativo	
Fluxo de Caixa Operacional	27
Total do Ativo	

Fonte: Pereira e Martins (2015)

Os indicadores Capital de Giro dividido pelo Total do Ativo e Lucros Acumulados sobre o Total do Ativo, apesar de terem aparecido em muitos trabalhos consultados, foram desconsiderados neste estudo. O primeiro foi devido à divergência de sua definição entre autores da área. Para Silva (2014), o capital de giro pode ser igual ao capital circulante líquido que é a diferença entre o ativo circulante e o passivo circulante. Para Gitman (2007), pode ser entendido como o ativo circulante, apenas. O segundo foi decorrente da Lei nº 11.638, de 28 de dezembro de 2007 que extingue a possibilidade de saldo da conta Lucros Acumulados. Dessa forma, o referido indicador teria resultado igual a zero em todo o período de análise.

Por tais motivos, apenas os dez indicadores restantes foram levados à diante na investigação. Por meio do método *stepwise* foram selecionadas as variáveis independentes que entrariam no modelo de regressão logística. Este método faz iterações, adicionando (*forward*) e removendo (*backward*) variáveis, de acordo com



um critério de seleção. Tais critérios podem ser o teste F, coeficiente de correlação linear múltipla, erro quadrático total etc (Hocking, 1976). Neste trabalho, foi utilizado o critério do teste F de Fisher. Levine et al (2008) comentam que este método verifica as variáveis que são significantes e as que não são para entrar no modelo.

Para os outros modelos, o método de seleção de variáveis escolhido foi o algoritmo genético. Primeiramente, inicia-se gerando uma população, em seguida seleciona indivíduos com melhores resultados de interesse e continua realizando iterações para gerar novos indivíduos, levando em consideração os indivíduos da população anterior. Esse processo se repete até encontrar o resultado ótimo (Coley, 1999). Ambos os métodos de seleção de variáveis foram implementados em MATLAB (Mathworks Inc., Natick, MA). Além deles, todos os classificadores foram executados no mesmo *software*, utilizando os mesmos grupos de treinamento e de teste, por meio dos indicadores financeiros.

Para os grupos de treinamento, foram utilizados os dados das empresas dos anos de 2010 a 2014 e 2011 a 2014, separadamente, uma vez que 2010 foi o ano de implementação plena das normas internacionais no Brasil, enquanto para os grupos de teste os de 2015 e 2016, para que fosse feita a validação. Como o número de companhias insolventes era menor do que o de solventes em todos os anos, apresentou-se um caso desbalanceado de classes. Visando a solução deste problema, selecionou-se aleatoriamente uma quantidade de empresas solventes igual ao número total de companhias insolventes para cada um dos anos (Chawla et al, 2002).

Foram utilizadas todas as empresas que apresentaram situação de insolvência em cada ano (2010 a 2014), sendo que foram todas agrupadas no grupo de treinamento, totalizando 132. Como havia 1.278 companhias em estado de solvência, no somatório do mesmo período, foram escolhidas 132 empresas, de maneira aleatória, para que fosse feito o balanceamento das classes. No ano de 2015 foram utilizadas todas as entidades que apresentaram estado de insolvência, totalizando 43. Neste mesmo ano, 226 evidenciaram solvência. Como também houve desbalanceamento de classes, foram selecionadas 43 das 226 organizações para compor o grupo de teste.



Em 2016, o total de companhias insolventes foi 41 e de solventes 260. Da mesma forma, como ocorreu desbalanceamento de classes em 2015, foram escolhidas as mesmas organizações solventes do ano 2015. As empresas utilizadas no estudo, tanto no grupo de treinamento quanto de teste estão dispostas no anexo. Na Tabela 2, estão presentes as variáveis selecionadas pelo *stepwise* para a RL e aquelas escolhidas pelo algoritmo genético para os classificadores LDA, MLP e SVM.

Tabela 2: Variáveis selecionadas para cada classificador

Variável	Indicador	Classificador			
		LDA	RL	MLP	SVM
X1	$\frac{\text{Lucro Líquido}}{\text{Total do Ativo}}$	X			X
X2	$\frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$	X	X		X
X3	$\frac{\text{EBIT}}{\text{Total do Ativo}}$		X		
X4	$\frac{\text{Vendas}}{\text{Total do Ativo}}$				X
X5	$\frac{(\text{Ativo Circulante} - \text{Estoques})}{\text{Passivo Circulante}}$	X	X		
X6	$\frac{\text{Dívida Total}}{\text{Total do Ativo}}$	X	X	X	X
X7	$\frac{\text{Ativo Circulante}}{\text{Total do Ativo}}$				
X8	$\frac{\text{Lucro Líquido}}{\text{Patrimônio Líquido}}$	X	X	X	X
X9	$\frac{\text{Caixa}}{\text{Total do Ativo}}$	X	X		
X10	$\frac{\text{Fluxo de Caixa Operacional}}{\text{Total do Ativo}}$				

Fonte: Dados da Pesquisa (2018)

Essas variáveis foram utilizadas para os classificadores análise discriminante, redes neurais e máquinas de vetores de suporte porque geraram um resultado superior em relação às variáveis escolhidas pelo *stepwise*.

3. ANÁLISE DOS RESULTADOS

A Tabela 3 e 4 mostram os resultados das acurácias obtidos por cada classificador em 2015, sendo a primeira (Tabela 3) referente ao grupo de treinamento de 2010 a 2014 e a segunda (Tabela 4) de 2011 a 2014. Já as Tabelas 5 e 6, apontam os valores resultantes das acurácias obtidas pelos classificadores em 2016, sendo que a cinco é



referente ao conjunto de treinamento de 2010 a 2014 e a seis ao íterim de 2011 a 2014. Em todos os casos foram utilizadas as variáveis selecionadas.

Tabela 3: Acurácia de cada classificador em 2015 (grupo de treinamento de 2010 a 2014)

Classificador	Acurácia (%)
Análise Discriminante	82,85
Regressão Logística	97,67
Redes Neurais	95,34
Máquinas de Vetores de Suporte	97,67

Fonte: Dados da pesquisa (2018)

Tabela 4: Acurácia de cada classificador em 2015 (grupo de treinamento de 2011 a 2014)

Classificador	Acurácia (%)
Análise Discriminante	87,20
Regressão Logística	98,83
Redes Neurais	90,69
Máquinas de Vetores de Suporte	90,69

Fonte: Dados da pesquisa (2018)

Tabela 5: Acurácia de cada classificador em 2016 (grupo de treinamento de 2010 a 2014)

Classificador	Acurácia (%)
Análise Discriminante	81,70
Regressão Logística	98,78
Redes Neurais	92,68
Máquinas de Vetores de Suporte	91,46



Fonte: Dados da pesquisa (2018)

Tabela 6: Acurácia de cada classificador em 2016 (grupo de treinamento de 2011 a 2014)

Classificador	Acurácia (%)
Análise Discriminante	80,48
Regressão Logística	98,78
Redes Neurais	89,02
Máquinas de Vetores de Suporte	92,68

Fonte: Dados da pesquisa (2018)

De maneira geral, o classificador que apresentou maior acurácia em ambos os anos, considerando os dois grupos de treinamento, foi a RL, para o conjunto de dados utilizados nesta pesquisa. O modelo com as menores acurácias em ambos os anos foi a análise discriminante, para os mesmos conjuntos de dados. Em 2015, de acordo com os dados apresentados na Tabela 3, a regressão logística não seria priorizada em relação às máquinas de vetores de suporte, já que estas, apesar de terem apresentado a mesma acurácia que aquela, utilizou uma variável a menos. Cabe ressaltar que as SVM foram utilizadas com o kernel rbf, devido ao melhor resultado extraído. Portanto, pelo critério da parcimônia, as SVM apresentariam um resultado superior.

Entretanto, ao considerar os resultados da Tabela 4, a acurácia das SVM cai – empatando com as MLP - e o da RL aumenta. Sendo assim, parece que a regressão logística, em 2015, foi o classificador mais acurado para os dados analisados, uma vez que praticamente empatou com as SVM no primeiro grupo de treinamento (Tabela 3) e se destacou no segundo (Tabela 4). Outra questão que merece ser levantada, é que as redes neurais artificiais obtiveram acurácias diferentes em 2015, devido à mudança do grupo de treinamento. Inclusive, as MLP e as máquinas de vetores de suporte apresentaram o mesmo resultado em 2015, considerando o grupo de treinamento de 2011 a 2014.



Sendo assim, não houve clareza em relação a qual dos modelos baseados em aprendizado de máquina se deve priorizar nesta pesquisa. O único classificador que apresentou a acurácia mais baixa de todas em 2015 foi a LDA, com 82,85% (Tabela 3) e 87,20% (Tabela 4). Contudo, em relação ao grupo de treinamento de 2011 a 2014, este modelo apresentou um aumento de acurácia de cerca de 5%, a RL incremento de aproximadamente 1%, as SVM e as MLP tiveram reduções por volta de 7% e 5%, respectivamente. Como os resultados do ano de 2015 podem gerar dúvidas em relação ao classificador mais indicado para análise de insolvência, acredita-se que tal fato pode ter sido relativo à mudança de grupos de treinamento, principalmente por causa do ano 2010.

Por conseguinte, em 2015, não houve unanimidade na escolha do melhor modelo, tendo em vista que em relação aos resultados da Tabela 3, as SVM indicaram um resultado um pouco melhor do que o da RL. Contudo, de acordo com a Tabela 4, a RL apresentou sozinha a maior acurácia. Em 2016, tanto para o conjunto de treinamento de 2010 a 2014 (Tabela 5) quanto para aquele de 2011 a 2014 (Tabela 6), a RL apresentou a mesma acurácia (98,78%), além de ter sido a maior dentre os classificadores utilizados nesse período. Entretanto, apesar das baixas variações das acurácias dos classificadores nos grupos de teste em 2016, exceto as da regressão logística que foram idênticas, as MLP e as SVM mudariam de posição no que tange a maior acurácia.

De acordo com os resultados da Tabela 5, as redes neurais apresentaram uma acurácia de 92,68%, sendo o classificador com a maior acurácia depois da regressão logística. Em seguida, seria as máquinas de vetores de suporte com 91,46%. Já em consideração aos resultados da Tabela 6, essa ordem foi invertida, ou seja, a segunda maior acurácia depois da RL foi a das SVM com 92,68%, seguida do referente às MLP com 89,02%. Em suma, mesmo com a pequena variação ocorrida nas acurácias dos classificadores baseados em aprendizado de máquina, os resultados poderiam desorientar os interessados na utilização de tais modelos. Os classificadores utilizados neste trabalho foram escolhidos por terem sido alvo de diversos estudos



sobre o tema e por terem mostrado uma boa capacidade preditiva, no que tange insolvência de empresas.

Por tais motivos, foram realizados testes para investigar qual dos modelos apresentaria melhor acurácia para esse fim. Para os resultados da acurácia de cada classificador, foram utilizados os indicadores apresentados na Tabela 2. As únicas variáveis consideradas nos quatro classificadores foram X6 e X8. A primeira representa quanto uma entidade possui de passivo exigível em relação ao total de ativos. Essa relação tem o intuito de evidenciar o montante de ativos financiados com capital de terceiros (Matarazzo, 2010). É possível, portanto, que esse indicador financeiro seja relevante na predição de insolvência, já que essa forma de financiamento de ativos pode gerar complicações financeiras para as companhias.

A variável X8 é a relação entre o lucro líquido sobre o patrimônio líquido, o que representa a rentabilidade do capital próprio. Sendo assim, é o retorno sobre o capital investido (Assaf, 2012). Dessa forma, também é possível que essa variável seja relevante para o estudo de previsão de insolvência de empresas. As variáveis X7 e X10 não foram selecionadas por nenhum dos modelos de seleção de variáveis junto aos classificadores. Acredita-se que esses indicadores não possuem aderência aos quatro métodos utilizados. Além disso, o primeiro apresenta endogenia, já que o indicador é a relação entre o ativo circulante e o ativo total, uma vez que o numerador está contido no denominador.

Portanto, parece que os modelos não foram afetados por essa potencial limitação, inclusive porque as acurácias foram relativamente elevadas. As acurácias apresentadas por cada modelo nos dois anos de validação não sofreram grandes alterações. Acredita-se, então, que as mudanças contábeis que vêm ocorrendo ao longo dos anos no Brasil não impactam significativamente na predição de insolvência das companhias listadas na Bolsa de Valores de São Paulo.



3.1 COMPARATIVO DE RESULTADOS DE ALGUNS TRABALHOS BRASILEIROS SEMELHANTES

Kanitz (1978) utilizou a análise discriminante como modelo para prever insolvência, considerando cinco indicadores financeiros como variáveis explicativas. A amostra selecionada pelo autor foi de 74 empresas brasileiras, sendo que não foi direcionada a um setor específico e continha 49 entidades solventes e 25 insolventes. A acurácia obtida pelo classificador foi de 86,62%. Altman *et al.* (1979) também utilizaram a análise discriminante e dados contábeis de empresas brasileiras. O total destas foi de 58, sendo 35 solventes e 23 insolventes. A acurácia obtida pelo modelo foi de 88%. Portanto, esses resultados podem sugerir que esse classificador apresenta uma capacidade um tanto quanto limitada de predição de insolvência de empresas, uma vez que não alcançou mais de 88% de acurácia, nessas pesquisas.

Horta *et al* (2011) utilizaram quatro classificadores: regressão logística (RL), redes neurais artificiais (MLP), máquinas de vetores de suporte (SVM) e árvore de decisão (AD). O classificador que apresentou a maior acurácia foi a AD com 99,3%. O período utilizado pelos autores foi de 1996 a 2006 e estes também utilizaram amostra genérica, ou seja, não analisaram companhias de setores econômicos específicos. Além disso, a pesquisa dos autores também foi baseada em empresas brasileiras listadas na Bolsa de Valores de São Paulo. Horta, Alves e Carvalho (2014) utilizaram dois classificadores: regressão logística e redes neurais artificiais (*backpropagation*), tendo obtido acurácias de 91,80% e 93,98%, respectivamente.

A amostra foi composta por 183 entidades brasileiras de setores diversos, desmembrada em 122 em situação solvente e 61 em estado de insolvência. O período analisado foi de 2004 a 2011. Soares e Rebouças (2014) realizaram um estudo com uma amostra de 87 companhias de capital aberto, divididas em 66 solventes e 21 insolventes, no período de 2002 a 2012. Cabe ressaltar que os dados contábeis extraídos pelos autores também foram oriundos do Economatica®, como nesta presente pesquisa. Soares e Rebouças (2014) utilizaram as seguintes técnicas: análise discriminante, regressão logística, árvores de classificação e redes neurais



artificiais. O modelo mais acurado foi as redes neurais artificiais com 98,85%, seguido das árvores de classificação com 94,25%, depois sendo a análise discriminante com 91,95% e, por último a regressão logística com 90,80%.

Horta *et al* (2015) utilizaram os classificadores regressão logística, árvore de decisão e máquinas de vetores de suporte. Além disso, a amostra também foi baseada em empresas brasileiras consideradas em estado de insolvência pelo SERASA e pela BOVESPA no intervalo de 2005 a 2007, não estando ligadas diretamente a um determinado setor. As acurácias obtidas pelos modelos foram 89,88% para a regressão logística; 93,45% para a árvore de decisão; e 97,02% para as máquinas de vetores de suporte. Prado (2016), em sua dissertação, realizou uma pesquisa sobre risco de crédito, a qual foi selecionada uma amostra de companhias, de diferentes setores, listadas na Bolsa de Valores de São Paulo, sendo que os dados contábeis foram extraídos do banco de dados do Economatica®.

A amostra teve a seguinte característica: 121 empresas no total, sendo que 70 apresentavam estado de solvência e 51 de insolvência. A delimitação temporal foi o período compreendido entre 1995 e 2014, inclusive. Os classificadores envolvidos na pesquisa foram a análise discriminante, regressão logística e redes neurais artificiais. Este último foi o modelo que apresentou a maior acurácia (97,8%) e os outros dois obtiveram o mesmo resultado (90,9%). Ao comparar os resultados desta pesquisa com os dos trabalhos citados nesta subseção, é possível observar que não há unanimidade em relação à qual é o classificador de previsão de insolvência que apresenta a maior acurácia. As pesquisas comparativas a este trabalho apresentam alguns pontos em comum e outros divergentes. De forma geral, A LDA não apresenta acurácias elevadas em relação aos outros classificadores.

Além disso, os métodos baseados em aprendizado de máquina são os que, geralmente, classificam com maior número de acerto as entidades em solventes e insolventes. Entretanto, os resultados desta pesquisa indicaram que o método que apresentou maior acurácia na predição de insolvência de companhias não financeiras listadas na BM&FBOVESPA foi a regressão logística. Entretanto, os resultados entre os quatro classificadores utilizados não foram tão distintos, exceto os da LDA. Outro



ponto divergente, é que os períodos de análise entre as pesquisas são diferentes, o que pode trazer diferenças entre os conjuntos de dados e, conseqüentemente, a classificação das entidades entre solventes e insolventes por parte dos classificadores.

Ao longo das últimas décadas, a análise discriminante foi o classificador de insolvência mais utilizado em trabalhos científicos (Pereira; Martins, 2015). Com o aprimoramento dos métodos quantitativos, em especial àqueles relacionados ao aprendizado de máquina, essa realidade vem mudando e apresentando resultados mais acurados do que os modelos mais antigos. Contudo, nesta pesquisa, no conjunto de dados de teste de 2015 e 2016, essa tendência não foi seguida, tendo em vista que a RL apresentou a maior acurácia. Cabe ressaltar que as redes neurais, a regressão logística e as máquinas de vetores de suporte são classificadoras não lineares.

Dessa forma, apesar de a LDA ter aparecido em diversos trabalhos, foi um classificador que apresentou baixa acurácia ao classificar empresas solventes e insolventes. Isso pode indicar que modelos lineares, neste trabalho, podem não ser os mais indicados. Outra questão relevante é que apesar de as companhias de diferentes setores terem formado as amostras das pesquisas ao longo dos anos, isso parece não ter afetado a classificação de insolvência de empresas, uma vez que as acurácias permaneceram relativamente elevadas. Tal fato pode ser observado nas obras dos autores citados nesta subseção, assim como na presente pesquisa.

CONSIDERAÇÕES FINAIS

Este trabalho teve o objetivo de investigar qual dos classificadores estudados – análise discriminante (LDA), regressão logística (RL), redes neurais (MLP) e máquinas de vetores de suporte (SVM) – apresentaria a melhor acurácia ao prever empresas insolventes, no período estudado. Dos quatro classificadores, apenas a análise discriminante apresentou divergência, no que tange a acurácia, já que os seus resultados foram inferiores aos demais nos dois anos de validação. As máquinas de vetores de suporte apresentaram o melhor desempenho em 2015, para o conjunto de



treinamento de 2010 a 2014, considerando o critério da parcimônia. Isso ocorreu por este modelo ter utilizado uma variável a menos do que a regressão logística, mesmo tendo obtido a mesma acurácia de 97,67%, ao classificar as companhias como insolventes.

Entretanto, a RL gerou uma acurácia maior do que todos os outros classificadores em 2015, considerando o grupo de treinamento de 2011 a 2014, acertando 98,83% da situação das companhias envolvidas. De forma geral, para os dados do ano de 2015, a RL se mostrou o classificador com as maiores acurácias para os dois grupos de treinamento, sendo 97,67% para o de 2010 a 2014 e 98,83% para o de 2011 a 2014. Em seguida, seria as SVM o segundo classificador com melhores resultados para os dados de 2015. Para o primeiro grupo de treinamento (2010 a 2014), esse modelo gerou uma acurácia de 97,67%, empatando com a RL em nível de percentual de acurácia. Já no segundo grupo de treinamento (2011 a 2014), as máquinas de vetores de suporte empataram com as redes neurais artificiais, com 90,69% de acurácia.

Como para o primeiro grupo as SVM apresentaram um resultado superior ao das MLP e no segundo estes classificadores empataram, as máquinas de vetores de suporte ficaram na segunda colocação. Na sequência, as redes neurais artificiais ficaram em terceiro lugar em 2015, tendo em vista que apresentaram a terceira melhor acurácia no primeiro grupo de treinamento (95,34%) e empatou com as SVM no segundo grupo (90,69%). A última colocação ficou com a LDA, uma vez que nos dois grupos de treinamento apresentou as menores acurácias, sendo 82,85% para o primeiro grupo (2010 a 2014) e 87,20% para o segundo (2011 a 2014). A regressão logística, apontou a maior acurácia (98,78%) em 2016, ao classificar as empresas em solventes e insolventes neste ano de teste, tanto em relação ao primeiro grupo de treinamento quanto ao segundo.

Na sequência, há uma indefinição em relação a qual classificador apontou melhores resultados. Isso ocorreu porque as redes neurais indicaram uma acurácia de 92,68% em relação ao primeiro grupo de treinamento, enquanto as máquinas de vetores de suporte 91,46%. Todavia, no que tange o segundo grupo, as MLP apresentaram uma redução de acurácia para 89,02% e as SVM um aumento para 92,68%. Sendo assim,



não há clareza em relação a qual modelo seria o mais acurado para prever insolvência com esses tipos de dados. A LDA assumiu o último lugar de priorização dos classificadores com 81,70% em relação ao primeiro grupo de treinamento e com 80,48% para o segundo. Utilizou-se dois anos para validação, com o intuito de averiguar se mudaria a classificação de acurácia dos classificadores, no que tange a predição de insolvência de companhias não financeiras listadas na BM&FBOVESPA.

Conforme os resultados expressos nas Tabelas 3, 4, 5 e 6, o modelo que se mostrou mais acurado em 2015 e 2016 foi a regressão logística, contrariando os resultados das pesquisas realizadas nos últimos anos. Devido às semelhanças de acurácias entre os classificadores, exceto as da LDA, e seus elevados índices de acerto ao classificar solvência e insolvência de companhias, as mudanças ocorridas nas IFRS parecem não ter prejudicado os resultados. Dentre as limitações deste trabalho, encontra-se a falta de evidências sobre quais indicadores financeiros contribuem para a previsão de insolvência. Cada um dos quatro classificadores estudados selecionou diferentes variáveis explicativas, tendo apenas duas em comum entre eles.

Além dessas questões, pode ter ocorrido um problema de endogenia na seleção das variáveis, como foi o caso da variável X9, que apresenta a relação entre a conta “caixa” e “total do ativo”, já que a primeira está contida na segunda. Contudo, pela relativa acurácia elevada dos classificadores, a endogenia talvez não tenha sido um empecilho para esta pesquisa. Por fim, como os indicadores financeiros não foram unânimes para todos os classificadores, sugere-se a utilização de outras variáveis explicativas como *inputs* e *outputs* para a classificação de empresas insolventes em trabalhos futuros. Além disso, outras formas de classificação de variáveis e outros modelos para atingir a mesma finalidade.

REFERÊNCIAS

Altman, E. I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. **Journal of Finance**, v. 23, n. 4, p. 589-609, 1968.



Altman, E. I., Baidya, T. K. N.; Dias. L. M. R. Previsão de problemas financeiros em empresas. **RAE: Revista de Administração de Empresas**, v. 19, n. 1, p. 17-28, 1979.

Assaf, A., Neto. **Estrutura e análise de balanços: um enfoque econômico-financeiro**. São Paulo: Atlas, 2012.

Beaver, W. H. Alternative accounting measures as predictors of failure. **The Accounting Review**, v. 43, n. 1, p.112-122, 1968.

Lei nº 11.638, de 28 de dezembro de 2007 (2007). Altera e revoga dispositivos da Lei nº 6.404, de 15 de dezembro de 1976. Diário Oficial da União. Brasília, DF.

Brito, G. A. S; Assaf, A. Modelo de classificação de risco de crédito de empresas. **Revista Contabilidade & Finanças**, USP – São Paulo, v. 19, n. 46, p. 18-29, 2008.

Chawla, N. V. et al. SMOTE: synthetic minority over-sampling technique. **Journal of Artificial Intelligence Research**, v. 16, n. 1, p. 321-357, 2002.

Coelho, F. U. **Curso de direito comercial: direito de empresa**. São Paulo: Saraiva, 2010.

Coley, D. A. **An introduction to genetic algorithms for scientists and engineers**. Singapore: World Scientific, 1999.

Cooper, D. R; Schindler, P. S. **Métodos de pesquisa em administração**. Porto Alegre: Bookman, 2003.

Cristianini, N; Taylor, J. S. **An introduction to support vector machines and other Kernel-based learning methods**. Cambridge: Cambridge University, 2003.

Deakin, E. B. A discriminant analysis of predictors of business failure. **Journal of Accounting Research**, v. 10, n. 1, p. 167-179, 1972.



Edmister, R. O. An empirical test of financial ratio analysis for small business failure prediction. **Journal of Financial and Quantitative Analysis**, v. 7, n. 2, p. 1477-1493, 1972.

Elizabetsky, R. **Um modelo matemático para decisões de crédito no banco comercial**. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, SP, Brasil, 1976.

Fávero, L. P. et al. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009.

Fitzpatrick, P. J. A comparison of ratios of successful industrial enterprises with those of failed Firms. **Certified Public Accountant**, p. 598-605, 1932.

Gitamn, L. J. **Princípios de administração financeira**. São Paulo: Pearson, 2007.

Hamel, L. **Knowledge discovery with support vector machines**. New Jersey: Wiley, 2009.

Haykin, S. **Neural networks: a comprehensive foundation**. New Jersey: Prentice Hall International, 1999.

Hocking, R. R. The analysis and selection of variables in linear regression. **Biometrics**, v. 32, n. 1, p. 1-49, 1976.

Horta, R. A. M., Borges, C. C. H., Carvalho F. A. A., & Alves, F. J. S. (2011). Previsão de insolvência: uma estratégia para balanceamento de base de dados utilizando variáveis contábeis de empresas brasileiras. *Sociedade, Contabilidade e Gestão*, UFRJ – Rio de Janeiro, 6(2), 21-36.

Horta, R. A. M. et al. Seleção de atributos na previsão de insolvência: aplicação e avaliação usando dados brasileiros recentes. **RAM: Revista de Administração Mackenzie**, v. 15, n. 1, p. 125-151, 2014.



Horta, R. A. M. et al. Seleção de atributos contábeis na previsão de insolvência de empresas brasileiras: uma comparação de abordagens. **BASE: Revista de Administração e Contabilidade da Unisinos**, v. 12, n. 1, p. 27-39, 2015.

Iudícibus, S. **Análise de balanços**. São Paulo: Atlas, 2009.

Kanitz, S. C. **Como prever falências**. São Paulo: Mc Graw-Hill do Brasil, 1978.

Levine, D. M. et al. **Estatística: teoria e aplicações**. Rio de Janeiro: LCT, 2008.

Lorena, A. C.; Carvalho, A. C. P. L. F. (2007). Uma introdução às support vector machines. **RITA: Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43-67.

Malhotra, N. K. *Pesquisa de marketing: uma orientação aplicada*. Porto Alegre: Bookman, 2001.

Matarazzo, D. C. **Análise de balanços: abordagem gerencial**. São Paulo: Atlas, 2010.

Matias, A. B. **Contribuição às técnicas de análise financeira**: um modelo de concessão de crédito. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, SP, Brasil, 1978.

Pasini, A. K. As normas internacionais de contabilidade – IFRS: adoção no Brasil. **RISUS: Journal on Innovation and Sustainability**, v. 6, n. 3, p. 97-114, 2015.

Pereira, V. S; Martins, V. F. Estudos de previsão de falências: uma revisão das publicações internacionais e brasileiras. **RCC: Revista Contemporânea de Contabilidade**, v. 12, n. 26, p. 163-196, 2015.

Prado, J. W. **Risco de crédito: uma abordagem utilizando análise discriminante, regressão logística e redes neurais artificiais**. Dissertação de Mestrado, Universidade Federal de Lavras, Minas Gerais, MG, Brasil, 2016.



ROSS, S. A; Westerfield, R. W; Jaffe, J. F. **Administração financeira: corporate finance**. São Paulo: Atlas, 2010.

Santos, E. S; Ponte, V. M. R; Mapurunga, P. V. R. Adoção obrigatória do IFRS no Brasil (2010): índice de conformidade das empresas com a divulgação requerida e alguns fatores explicativos. **Revista Contabilidade & Finanças**, v. 25, n. 65, p. 161-176, 2014.

Schmueli, G; Patel, N. R; Bruce, P. C. **Data Mining for Business Intelligence: concepts, techniques and applications in Microsoft Office Excel with XLMiner**. New Jersey: Wiley, 2010.

Soares, R. A; Rebouças, S. M. D. P. Avaliação do desempenho de técnicas de classificação aplicadas à previsão de insolvência de empresas de capital aberto brasileiras. **Revista ADM.MADE**, v. 18, n. 3, p. 41-61, 2014.

Silva, A. A. **Estrutura, análise e interpretação das demonstrações contábeis**. São Paulo: Atlas, 2014.

ANEXO

Empresas insolventes em 2010

Refinaria de Petróleos Manguinhos; Sansuy Indústria de Plásticos; Recrusul; Nordon Indústrias Metalúrgicas; Minupar Participações; Laep Investments; Bombril; Agrenco; Construtora Adolpho Lindenberg; Tec Blumenau; Fiação e Tecelagem São José; Teka-Tecelagem Kuehnrich; Têxtil Renauxview; IGB Eletrônica; Hércules Fábrica de Talheres; DHB Indústria e Comércio; Hotéis Othon; Sauipe; Manufatura de Brinquedos Estrela; Biomm; Celg de Participações; Sanesalto Saneamento.

Empresas solventes em 2010

Cosan Indústria e Comércio; Óleo e Gás Participações; Petro Rio; QGEP Participações; Ultrapar Participações; Lupatech; OSX Brasil; MMX Mineração e Metálicos; Paranapanema; Fertilizantes Hering; Duratex; Fibria Celulose;



Magnesita Refratários; Eternit; PBG; Mills Estruturas e Serviços de Engenharia; Embraer; Tupy; WEG; Indústrias Romi; Metalfrio Solutions; LOG-IN Logística Intermodal.

Empresas insolventes em 2011

Lupatech; Refinaria de Petróleos Manguinhos; Sansuy Indústria de Plásticos; Haga Indústria e Comércio; Recrusul; Nordon Indústrias Metalúrgicas; Advanced Digital Health Medicina Preventiva; Bombril; Agrenco; Construtora Adolpho Lindenberg; Buettner Indústria e Comércio; Tec Blumenau; Fiação e Tecelagem São José; Teka-Tecelagem Kuehnrich; Têxtil Renauxview; IGB Eletrônica; Hércules Fábrica de Talheres; DHB Indústria e Comércio; Hotéis Othon; Sauipe; Manufatura de Brinquedos Estrela; Biomm; Celg de Participações; Sanesalto Saneamento.

Empresas solventes em 2011

Cosan Indústria e Comércio; Óleo e Gás Participações; Petro Rio; QGEP Participações; Ultrapar Participações; OSX Brasil; MMX Mineração e Metálicos; Paranapanema; Fertilizantes Hering; Duratex; Fibria Celulose; Magnesita Refratários; Eternit; PBG; Mills Estruturas e Serviços de Engenharia; Embraer; Tupy; WEG; Indústrias Romi; Metalfrio Solutions; LOG-IN Logística Intermodal; JSL; TGMA Gestão Logística; CCR.

Empresas insolventes em 2012

Lupatech; Refinaria de Petróleos Manguinhos; Sansuy Indústria de Plásticos; Haga Indústria e Comércio; Manufatura de Brinquedos Estrela; Biomm; Celg de Participações; Rede Energia; Sanesalto Saneamento; Minupar Participações; Laep Investments; Advanced Digital Health Medicina Preventiva; Bombril; Agrenco; Construtora Adolpho Lindenberg; Buettner Indústria e Comércio; Tec Blumenau; Fiação e Tecelagem São José; Teka-Tecelagem Kuehnrich; Têxtil Renauxview; IGB Eletrônica; Hércules Fábrica de Talheres; DHB Indústria e Comércio; Hotéis Othon; Sauipe; Recrusul; Wetzel; Nordon Indústrias Metalúrgicas.

Empresas solventes em 2012



Cosan Indústria e Comércio; Óleo e Gás Participações; Petro Rio; QGEP Participações; Ultrapar Participações; OSX Brasil; MMX Mineração e Metálicos; CCX Carvão da Colômbia; Paranapanema; Fertilizantes Hering; Duratex; Fibria Celulose; Magnesita Refratários; Eternit; PBG; Mills Estruturas e Serviços de Engenharia; Embraer; Tupy; WEG; Indústrias Romi; Metalfrio Solutions; LOG-IN Logística Intermodal; JSL; Tegma Gestão Logística; CCR; Ecorodovias Infraestrutura e Logística; Triunfo Particip. E Invest.; Prumo Logística.

Empresas insolventes em 2013

Óleo e Gás Participações; Lupatech; Rede Energia; Sanesalto Saneamento; OGX Petróleo e Gás; Refinaria de Petróleos Manguinhos; Mangels Industrial; Sansuy Indústria de Plásticos; Haga Indústria e Comércio; Recrusul; Wetzel; Nordon Indústrias Metalúrgicas; Minupar Participações; Advanced Digital Health Medicina Preventiva ; Bombril ; Buettner Indústrias e Comércio; Karsten; Tec Blumenau; Fiação e Tecelagem São José; Teka-Tecelagem Kuehnrich; Têxtil Renauxview; IGB Eletrônica; Hércules Fábrica de Talheres; DHB Indústria e Comércio; Hotéis Othon; Manufatura de Brinquedos Estrela; Hopi Hari; Celg de Participações.

Empresas solventes em 2013

Cosan Indústria e Comércio; Petro Rio; QGEP Participações; Ultrapar Participações; OSX Brasil; MMX Mineração e Metálicos; CCX Carvão da Colômbia; Paranapanema; Fertilizantes Hering; Duratex; Fibria Celulose; Magnesita Refratários; Eternit; PBG; Mills Estruturas e Serviços de Engenharia; Embraer; Tupy; WEG; Indústrias Romi; Metalfrio Solutions; Cosan Logística; Rumo Logística Operadora Multimodal; Logística Intermodal; JSL; Tegma Gestão Logística; CCR; Ecorodovias Infraestrutura e Logística; Triunfo Particip. E Invest.

Empresas insolventes em 2014

OSX Brasil; MMX Mineração e Metálicos; Gol Linhas Aéreas Inteligentes; Refinaria de Petróleos Manguinhos; Mangels Industrial; IGB Eletrônica; Hércules Fábrica de Talheres; Hotéis Othon; Manufatura de Brinquedos



Estrela; Tec Toy; Hopi Hari; Sanesalto Saneamento; Sansuy Indústria de Plásticos; Haga Indústria e Comércio; Cobrasma; Recrusul; Wetzel; Inepar Indústria e Construções; Nordon Indústrias Metalúrgicas; Minupar Participações; Advanced Digital Health Medicina Preventiva; Bombril; Buettner Indústria e Comércio; Karsten; Companhia Industrial Schlösser; Tec Blumenau; Fiação e Tecelagem São José; Teka-Tecelagem Kuehnrich; Têxtil Renauxview; Celg de Participações.

Empresas solventes em 2014

Prumo Logística; Contax Participações; Valid Soluções e Serv. Seg. Meios Pag. Ident.; Bras. de Prop. Agrícolas; Pomifrutas; SLC Agrícola; Terra Santa Agro; Biosev; São Martinho; BRF; JBS; Marfrig Global Foods; Minerva; M. Dias Branco Ind. Com. de Alimentos; Natura Cosméticos; Hypermarchas; CR2 Empreendimentos Imobiliários; Cyrela Brazil Realty Empreend. e Part.; Direcional Engenharia; Even Construtora e Incorporadora; Eztec Empreendimentos e Participações; Gafisa; Helbor Empreendimentos; JHSF Participações; MRV Engenharia e Participações; PDG Realty Empreend. e Participações; Rodobens Negócios Imobiliários; Rossi Residencial; Tecnisa; Trisul.

Empresas insolventes em 2015

OSX Brasil; MMX Mineração e Metálicos; Minerva; Viver Incorporadora e Construtora; Springs Global Participações; Ideiasnet; Estadual de Distrib. Ener. Elet.-Ceee-D; Forjas Taurus; Gol Linhas Aéreas Inteligentes; OGX Petróleo e Gás; Refinaria de Petróleos Manguinhos; Mangels Industrial; GPC Participações; Sansuy Indústria de Plásticos; Haga Indústria e Comércio; Sanesalto Saneamento; Cobrasma; Recrusul; Metalúrgica Riosulense; Wetzel; Inepar Indústria e Construções; Nordon Indústrias Metalúrgicas; Numeral 80 Participações; Minupar Participações; Advanced Digital Health Medicina Preventiva; Bombril; Buettner Indústria e Comércio; Karsten; Cia. Industrial Schlosser; Tec Blumenau; Fiação e Tecelagem São José; Teka-Tecelagem Kuehnrich; Têxtil Renauxview; Cambuci; Mundial – Produtos de Consumo; IGB Eletrônica; Hércules Fábrica de Talheres; Hotéis Othon; Manufatura de



Brinquedos Estrela; Tec Toy; Hopi Hari; Telec. Brasileiras Telebrás; Celg de Participações.

Empresas solventes em 2015

Cosan Indústria e Comércio; Petro Rio; QGEP Participações; Ultrapar Participações; Lupatech; CCX Carvão da Colômbia; Paranapanema; Fertilizantes Heringer; Duratex; Fibria Celulose; Magnesita Refratários; Eternit; PBG; Mills Estruturas e Serviços de Engenharia; Embraer; Tupy; WEG; Indústrias Romi; Metalfrio Solutions; Cosan Logística; Hering; Grendene; Technos; Unicasa Indústria de Móveis; Iochpe Maxion; Somos Educação; International Meal Company Alimentação; T4F Entretenimento; CVC Brasil Operadora e Agência de Viagens; Gaec Educação; Estácio Participações; Kroton Educacional; Ser Educacional; Localiza Rent a Car; Locação das Américas; Multiplus; Smiles; Arezzo Indústria e Comércio; Restoque Comércio e Confeções de Roupas; Marisa Lojas; Lojas Renner; Magazine Luiza; B2W – Companhia Digital.

Empresas insolventes em 2016

OSX Brasil; MMX Mineração e Metálicos; Viver Incorporadora e Construtora; Estadual de Distrib. Ener. Elet.-Ceee-D; Forjas Taurus; Gol Linhas Aéreas Inteligentes; OGX Petróleo e Gás; Refinaria de Petróleos Manguinhos; Mangels Industrial; Sansuy Indústria de Plásticos; Haga Indústria e Comércio; Sanesalto Saneamento; Recrusul; Metalúrgica Riosulense; Wetzel; Nordon Indústrias Metalúrgicas; Minupar Participações; Advanced Digital Health Medicina Preventiva; Bombril; Karsten; Teka-Tecelagem Kuehnrich; Têxtil Renauxview; Cambuci; Mundial – Produtos de Consumo; IGB Eletrônica; Hércules Fábrica de Talheres; Hotéis Othon; Manufatura de Brinquedos Estrela; Tec Toy; Telec. Brasileiras Telebrás; Celg de Participações; CCX Carvão da Colômbia; Mendes Junior Engenharia; Altus Sistema de Automação; Conc. Aeroporto Internacional Guarulhos; Plascar Participações Indust.; BR Home Centers; PDG Realty; Pomifrut; Dommo Energia; Fibam Companhia Industrial.

Empresas solventes em 2016



Cosan Indústria e Comércio; Petro Rio; QGEP Participações; Ultrapar Participações; Lupatech; Paranapanema; Fertilizantes Heringer; Duratex; Fibria Celulose; Magnesita Refratários; PBG; Mills Estruturas e Serviços de Engenharia; Embraer; Tupy; WEG; Indústrias Romi; Metalfrio Solutions; Cosan Logística; Hering; Grendene; Technos; Unicasa Indústria de Móveis; Iochpe Maxion; Somos Educação; International Meal Company Alimentação; T4F Entretenimento; CVC Brasil Operadora e Agência de Viagens; Gaec Educação; Estácio Participações; Kroton Educacional; Ser Educacional; Localiza Rent a Car; Locação das Américas; Multiplus; Smiles; Arezzo Indústria e Comércio; Restoque Comércio e Confecções de Roupas; Marisa Lojas; Lojas Renner; Magazine Luiza; B2W – Companhia Digital.

Enviado: Abril, 2019.

Aprovado: Agosto, 2019.