

CLASSIFICAÇÃO DE RISCO DE CRÉDITO UTILIZANDO ANÁLISE DISCRIMINANTE, REGRESSÃO LOGÍSTICA E REDES NEURAS ARTIFICIAIS

JOSÉ WILLER DO PRADO

UNIVERSIDADE FEDERAL DE LAVRAS (UFLA)
jwprado@gmail.com

FRANCISVAL DE MELO CARVALHO

UNIVERSIDADE FEDERAL DE LAVRAS (UFLA)
francarv@dae.ufla.br

ANDRÉ LUIS RIBEIRO LIMA

UNIVERSIDADE FEDERAL DE LAVRAS (UFLA)
andre.lima@dae.ufla.br

LUÍS FERNANDO SILVA ANDRADE

UNIVERSIDADE FEDERAL DE LAVRAS (UFLA)
andradelfs@gmail.com

Introdução

O risco está associado à probabilidade de que ocorra determinado resultado em relação ao retorno esperado. Está voltado para o futuro, revelando uma possibilidade de perda (DAMODARAN, 2011). Nas atividades financeiras de crédito busca-se transformar a incerteza em risco mensurável. Existem diversas ferramentas que se propõem à análise do crédito. As mais tradicionais, são as técnicas de análise multivariada. Contudo, técnicas de Inteligência artificial têm despertado a atenção dos pesquisadores.

Problema de Pesquisa e Objetivo

Martins (2003) afirma que a literatura não determina quais indicadores são mais expressivos na avaliação da insolvência. Neste sentido, o problema constitui-se em quais indicadores econômico-financeiros contribuem para uma melhor acurácia na análise da concessão de crédito para empresas brasileiras de capital aberto? Adiante ainda, aferir quais das técnicas utilizadas (análise discriminante, regressão logística e redes neurais) apresentam melhor acurácia para prever a falência empresarial.

Fundamentação Teórica

Segundo Sicsú (2010), quantificar o risco de crédito de forma objetiva, por meio de técnicas quantitativas de análise, tem apresentado grandes vantagens como, por exemplo, rapidez nas decisões, decisões mais consistentes, decisões adequadas e outras. Complementando, Kimura et al. (2008) afirmam que a gestão de risco, por meio da pesquisa e implementação de modelos quantitativos, tem se tornado uma das principais áreas dentro das empresas modernas.

Metodologia

Com base na literatura foram selecionados 35 indicadores. Por meio da Lei n. 7.661, e da Lei nº 11.101, foram escolhidas na base de dados Econômica 51 empresas insolventes e também 70 empresas solventes do mesmo período e setor para equiparar a amostra. Para a criação dos modelos, são utilizados análise discriminante, regressão logística e redes neurais artificiais. Para avaliar os modelos utilizou-se a Curva ROC (Receiver Operating Characteristics) e o teste de Kolmogorov-Smirnov – KS.

Análise dos Resultados

Todos os pressupostos das modelagens foram atendidos. Ao se comparar as técnicas utilizadas para prever a insolvência empresarial, por meio da Curva ROC (Receiver Operating Characteristics) do teste de Kolmogorov-Smirnov – KS, assim como, por meio do Erro tipo I e da acurácia de acerto, pode-se perceber a superioridade das redes neurais sobre a regressão logística e desta sobre a análise discriminante. Dos 35 indicadores selecionados, apenas 17 participaram dos agrupamentos finais para análise.

Conclusão

Para os três modelos e os sete agrupamentos de indicadores propostos, foram gerados 21 modelos, os quais apresentaram acurácia acima de 85% nas classificações, o que demonstra que os agrupamentos propostos obtiveram bons resultados. Acredita-se que, por meio do presente estudo, pode-se elucidar algumas das características que possuem as empresas insolventes da amostra. Tais contribuições são fundamentais para as pesquisas de risco de crédito e contribuem para o desenvolvimento da temática.

Referências Bibliográficas

- SICSÚ, A. L. Credit scoring: desenvolvimento, implantação, acompanhamento. SP: Blucher, 2010.
- MARTINS, M. S. A previsão de insolvência pelo modelo de Cox: uma contribuição para a análise de companhias abertas brasileiras. 2003.
- DAMODARAN, A. Investment Valuation: Tools and Techniques for Determining the Value of Any Asset. NY: John Wiley & Sons: 2011.
- KIMURA, H. et al. Value-at-Risk: como entender e calcular o risco pelo VaR: uma contribuição para a gestão no Brasil. RP: Inside Books, 2008.

Área temática: Finanças – Crédito e Análise Setorial.

CLASSIFICAÇÃO DE RISCO DE CRÉDITO UTILIZANDO ANÁLISE DISCRIMINANTE, REGRESSÃO LOGÍSTICA E REDES NEURAIS ARTIFICIAIS

1 INTRODUÇÃO

O risco está associado à probabilidade de que ocorra determinado resultado em relação ao retorno esperado. Assim, está voltado para o futuro, revelando uma possibilidade de perda (ASSAF NETO, 2010; DAMODARAN, 2011). Nas atividades financeiras envolvendo crédito busca-se encontrar segurança contra o risco presente nas operações ou, no mínimo, transformar a incerteza em risco mensurável (SILVA, 1983). Evidencia-se que as primeiras pesquisas neste campo buscavam detectar se os indicadores das empresas solventes eram favoráveis e se os indicadores das empresas insolventes eram desfavoráveis. Um dos primeiros estudos (univariados), realizados no campo, foram *The Use of Multiple Measurements in Taxonomic Problems* de Fisher (1936) e *Risk Elements in Consumer Installment Lending* de Durand (1941). Entretanto as análises univariadas, realizadas no final da década de 1950, foram substituídas assim que as pesquisas acadêmicas se voltaram para técnicas de modelagem de *credit scoring* no final da década de 1960 (SABATO, 2009).

Os trabalhos seminais neste campo foram o *Financial ratios predictors of failure* de Beaver (1966) e *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy* de Altman (1968). Esses dois estudos iniciaram as pesquisas com a *multivariate discriminant analysis*, técnica esta que viria a ser bastante utilizada nas pesquisas sobre a temática até Ohlson (1980) realizar um estudo chamado *Financial ratios and the probabilistic prediction of bankruptcy*, utilizando o modelo *logit* para a previsão de inadimplência, abrindo uma perspectiva para a utilização de outras metodologias no campo de estudo. Uma década depois, Odom e Sharda (1990), com o trabalho *A Neural Network Model for Bankruptcy Prediction*, propuseram um novo modelo, desta vez fundamentado nas técnicas de inteligência artificial conhecidas como *artificial neural networks*.

No Brasil, o primeiro estudo foi o artigo “Como prever falências de empresas”, publicado pelo professor Stephen Charles Kanitz e que utiliza análise discriminante e regressão múltipla. Kanitz (1974) propôs um termômetro da situação de solvência empresarial que se tornaria referência para pesquisas futuras. Atualmente existem diversas ferramentas que se propõem à análise do crédito. As mais tradicionais, na construção de modelos de risco de crédito, são as técnicas de análise multivariada como, por exemplo, a análise discriminante e a regressão logística. Contudo, técnicas de Inteligência Computacional têm despertado a atenção dos pesquisadores para novas possibilidades (CASTRO JUNIOR, 2003; CHUANG; HUANG, 2011; KIM, 2011; PRADO et al., 2016).

Martins (2003) afirma que a literatura financeira não determina quais os indicadores são mais expressivos na avaliação da insolvência. Segundo o autor, mesmo que muitos indicadores sejam empregados de forma recorrente em vários estudos, a escolha dos indicadores, em grande parte, está relacionada ao acesso aos dados e à percepção do pesquisador. Neste sentido, o **problema de pesquisa** constitui-se em quais indicadores econômico-financeiros contribuem para uma melhor acurácia na análise da concessão de crédito para empresas brasileiras de capital aberto? Adiante ainda, aferir quais das técnicas utilizadas (análise discriminante, regressão logística e redes neurais artificiais) apresentam melhor acurácia para prever a falência empresarial.

De uma perspectiva acadêmica, o estudo da solvência das empresas é importante, para o fortalecimento da linha de pesquisa em finanças, dada uma melhor compreensão do

comportamento das empresas e suas estratégias de financiamento. Por meio dessa proposta de pesquisa, pretende-se contribuir com a ampliação de conhecimentos a respeito da análise teórica e empírica sobre risco de crédito e sua potencial utilidade para o aprimoramento das decisões empresariais no mercado de capital no Brasil.

2 REFERENCIAL TEÓRICO

Soares e Rebouças (2015, p. 41) destacam que “uma das aplicações mais conhecidas da previsão de insolvência é auxiliar na redução do risco de crédito”. Segundo Assaf Neto e Silva (2012), crédito diz respeito à troca de bens presentes por bens futuros. Os autores complementam que, quando uma empresa concede crédito, troca produtos por uma promessa de pagamento futuro. Por outro lado, uma empresa que obtém crédito recebe produtos e assume o compromisso de efetuar o pagamento no futuro.

O crédito tem importante função econômica e social na vida das empresas. Silva (2008) afirma que o papel do crédito é possibilitar que as empresas aumentem seu nível de atividade, estimulem o consumo e o aumento da demanda e financiar a execução de projetos para os quais as empresas não disponham de recursos próprios suficientes. Contudo, cabe ressaltar que o uso inadequado do crédito pode levar uma empresa à falência (insolvência). Kimura et al. (2008, p. 16) observam, ainda, que:

Uma relação de crédito liga o credor ao devedor e estabelece uma obrigação para o devedor efetuar um determinado pagamento ao credor em uma data futura. São exemplos de operações que envolvem risco de crédito: o mútuo, as debêntures, CDB, RDB etc. Pragmaticamente, qualquer empréstimo de recursos para posterior recebimento implica em um risco de crédito.

Em seu estudo seminal na década de 1920, Knight (2002, p. 224) diferenciou risco de incerteza, definindo-a como “um estado em que não existem bases válidas ou experiência passada para determinar a probabilidade de uma ocorrência específica”. O autor associou o lucro à existência de incerteza.

No que tange à concessão de crédito, a incerteza é inerente a essa atividade e sempre estará presente. Para Sicsú (2010), as instituições financeiras sempre estarão sujeitas a possibilidades de perdas neste tipo de transação. Sendo assim, o recomendável seria quantificar a probabilidade de perda, possibilitando uma melhor classificação dos clientes a empréstimos. Segundo o autor, “à probabilidade de perda em uma operação de crédito denominamos risco de crédito” (SICSÚ, 2010, p. 1).

2.1 Risco de crédito

Em razão do fato de envolver uma expectativa de retorno do patrimônio que foi cedido, Schrickel (2000) explana que todo crédito está vinculado, de alguma forma, ao risco da perda do valor envolvido. Yu (2014) ressalta que avaliação do risco de crédito tem sido alvo de uma série de estudos em profundidade nos últimos anos, é o principal foco das áreas financeiras e bancárias em razão, principalmente, da recente crise financeira internacional, que teve um efeito grave em muitas organizações financeiras. Oreski e Oreski (2014) frisam que o risco de crédito é uma das questões mais importantes para o setor bancário e tem ganhado cada vez mais atenção.

Desta forma, Assaf Neto (2010) afirma que o risco é um custo que está sempre presente nos negócios e, portanto, deve ser quantificado. Assim Garcia, Gimenez e Guijarro (2013) elucidam que a gestão do risco de crédito é uma questão chave para qualquer empresa em qualquer momento. Os autores observam, ainda, que existem atualmente várias metodologias que visam prever a probabilidade de inadimplência dos devedores, muitas delas utilizando análise logit ou análise discriminante para classificá-los. Entretanto, novas metodologias fazem uso de redes neurais ou métodos multicritérios.

Harris (2013) observa que o aumento da concorrência, no setor de serviços financeiros, tem levado muitas empresas a encontrarem formas inovadoras para lidar com o risco, a fim de atingir e/ou manter uma vantagem competitiva. Como resultado do ambiente econômico e de negócio atual, as instituições financeiras enfrentam maior risco de perdas associadas com a inadequação de aprovação de crédito nas decisões. Entretanto, Kou e Wu (2014) afirmam que o principal objetivo da análise de risco de crédito é a de classificar as amostras de clientes em bons ou maus pagadores (solventes ou insolventes).

Segundo Sicsú (2010), quantificar o risco de crédito de forma objetiva, por meio de técnicas quantitativas de análise, tem apresentado grandes vantagens como, por exemplo, rapidez nas decisões, decisões mais consistentes, decisões adequadas e outras. Complementando, Kimura et al. (2008) afirmam que a gestão de risco, por meio da pesquisa e implementação de modelos quantitativos, tem se tornado uma das principais áreas dentro das empresas modernas.

Silva (2008) ressalta que o uso de métodos quantitativos tem sido muito divulgado, o que tem levado muitos bancos a uma corrida em busca dessas técnicas para avaliação do risco de crédito dos clientes devido à sua contribuição para o processo de análise. Entretanto, esses métodos de previsão não podem ser interpretados como ‘receitas milagrosas’ capazes de resolver todos os problemas advindos do risco de crédito contido nas operações.

3 METODOLOGIA

A pesquisa quanto aos fins, pode ser classificada como explicativa (VERGARA, 2008). No que tange aos meios, pode ser caracterizada como *ex post facto* (VERGARA, 2008). Quanto aos tipos de dados, como secundários, pesquisas bibliográfica e documental (MARCONI; LAKATOS, 2005), sendo a pesquisa documental realizada nas demonstrações financeiras anuais, disponíveis na base de dados Economática[®] e no site da BM&FBOVESPA (Bolsa de Valores, Mercadorias e Futuros). Quanto à forma de abordagem, esta pesquisa é qualificada como quantitativa (MARTINS; THEÓPHILO, 2009). Para o desenvolvimento do trabalho, utilizou-se os *softwares* SPSS (*Statistical Package for Social Sciences*) e Minitab.

Para coleta das demonstrações contábeis das empresas insolventes, é preciso, primeiramente, definir o conceito de solvência utilizado no estudo. Para definição de insolvência, foi utilizada a lei de falência e concordatas: Lei n. 7.661, de 21 de junho de 1945 (BRASIL, 1945), que foi revogada pela Lei n° 11.101, de fevereiro de 2005 (BRASIL, 2005), que passa a valer para todos os casos de falência e concordata atuais. Ou seja, o evento que determinará a entrada de uma empresa no estado de insolvência será apontado pela data em que a empresa tenha entrado com o pedido formal de Concordata Preventiva ou de Recuperação Judicial junto à Comissão de Valores Mobiliários (CVM) no período de 1995 a 2014. Como salienta Sanvicente e Minardi (1998, p. 1), a “previsão de concordatas é um aspecto essencial de modelos de gerenciamento de risco de crédito”.

Sendo assim, para a coleta dos indicadores das empresas insolventes, foi utilizada a data de um ano antes da empresa anunciar a falência (ano anterior do evento, tempo_{t-1}). Contudo, para se completar a amostra, seleciona-se, ao menos, uma empresa solvente para cada empresa insolvente, isto é, para cada empresa determinada insolvente, selecionar-se-á, no mínimo, uma empresa solvente pertencente ao mesmo setor e, quando possível, com ativos contábeis proporcionais ao da instituição insolvente. Essa metodologia baseia-se em estudos anteriores desenvolvidos por Altman (1968), Brito, Assaf Neto e Corrar (2009) e Sanvicente e Minardi (1998).

A lista com todas as empresas selecionadas pode ser visualizada no Quadro 1. São apresentadas as 51 empresas insolventes selecionadas juntamente com a data do seu anúncio de pedido de Concordata Preventiva ou de Recuperação Judicial, além do seu setor econômico e algumas das 70 empresa solvente que foram utilizadas para equiparar a amostra.

Quadro 1 Amostra de Empresas solventes e insolventes

Setor Econômico	Ano do anúncio	Nº	Empresa Insolvente	Nº	Empresa Solvente
Química	1995	1	Aquatec	52	Polialden
Outros / Administração	1995	2	Montreal	54	Trevisa
Comércio	1995	3	Mesbla	55	Lojas Renner
Construção	1996	4	Sv Engenhar	56	Mendes Jr
Outros / Curtumes	1997	5	Corbetta	57	Cedro
Veículos e peças	1997	6	Spscs Industrial	58	Randon Part
Comércio	1998	7	Lojas Arapua	59	Bompreco
Química	1998	8	Copas	60	Fertibras
Eletrônicos	2000	9	Sharp	61	Electrolux
Eletrônicos	2002	10	Inbrac	62	Trafo
Outros / Ind. Madeira	2003	11	Eucatex	64	Fibria
Alimentos e Bebidas	2004	12	Chapeco	66	Iguacu Cafe
Alimentos e Bebidas	2004	13	Parmalat	67	Leco
Transporte e Serviços	2004	14	Vasp	68	Tam S/A
Transporte e Serviços	2005	15	Varig Transp	69	Gol
Química	2005	16	Bombril	70	Pronor
Outros/Ind. Produtos de plást.	2005	17	Sansuy	71	Evora
Veículos e peças	2006	18	Recrusul	73	Wetzel S/A
Siderurgia & Metalurgia	2007	19	Kepler Weber	75	Tekno
Alimentos e Bebidas	2007	20	Minupar	76	Josapar
Têxtil	2008	21	Botucatu Tex	77	Karsten
Agroindústria e Pesca	2008	22	Agrenco	79	SLC Agrícola
Construção	2008	23	Const Beter	82	Azevedo
Petróleo e Gás	2008	24	Pet Manguinh	83	Ceg
Minerais não Metálicos	2008	25	Chiarelli	84	Eternit
Eletrônicos	2009	26	IGB S/A	85	Itautec
Têxtil	2010	27	Tex Renaux	86	Le Lis Blanc
Têxtil	2010	28	Tecel S Jose	88	Marisol
Têxtil	2011	29	Schlosser	89	Ind Cataguas
Têxtil	2011	30	Buettner	91	Cremer
Têxtil	2011	31	Fab C Renaux	94	Vicunha Text
Energia Elétrica	2012	32	Celpe	96	Celpe
Outros / Comercio	2012	33	Lark Maqs	99	Minasmaquinas
Outros / Consultoria	2012	34	Tecnosolo	100	Sondotecnica
Têxtil	2012	35	Teka	102	Springs
Energia Elétrica	2012	36	Rede Energia	104	Tractebel
Outros / Administração	2013	37	Gp Invest	106	Kroton
Outros / Administração	2013	38	GPC Part	107	Habitasul
Alimentos e Bebidas	2013	39	Clarion	108	Cacique
Petróleo e Gás	2013	40	OGX Petroleo	109	Petrobras
Siderurgia & Metalurgia	2013	41	Mangels Indl	110	Forja Taurus
Veículos e peças	2013	42	OSX Brasil	111	Tupy
Siderurgia & Metalurgia	2014	43	Met Duque	112	Aliperti
Siderurgia & Metalurgia	2014	44	Lupatech	113	Ferbasa
Outros / Outras indústrias	2014	45	Inepar	114	V-Agro
Siderurgia & Metalurgia	2014	46	Fibam	115	Metal Iguacu
Mineração	2014	47	MMX Miner	116	Magnesita SA
Energia Elétrica	2014	48	Eneva	117	AES Elpa
Veículos e peças	2015	49	DHB	118	Plascar Part
Outros / Administração	2015	50	Ideiasnet S.A.	120	Itausa
Construção	2015	51	Sultepa	121	Lix da Cunha

Fonte: Elaborado pelos autores. Obs.: 20 das 70 empresas solventes utilizadas no estudo foram omitidas do quadro para facilitar a visualização.

No que se refere à amplitude da amostra (tamanho), buscou-se um número de empresas insolventes que possibilitasse as análises de forma a não prejudicar o processo de modelagem para as técnicas abordadas. Neste sentido, para a elaboração do presente estudo, a amostra foi composta por 121 empresas, sendo 70 empresas consideradas solventes e 51 empresas insolventes – atendendo o recomendado por Hair et al. (2009, p. 236) de que deve haver no “mínimo 40 empresas” pertencentes aos dois grupos definidos a priori.

3.1 Definição dos indicadores

Para a seleção dos indicadores do Modelo Tradicional de análise financeira, foram utilizados os indicadores que mais são utilizados no levantamento realizado com trabalhos da área. A definição dos indicadores está fundamentada nas considerações de Pereira, Domínguez e Oejo (2007), em que os autores afirmam que a evidência empírica tem apontado que a escolha de indicadores que apresentaram resultados satisfatórios em pesquisas anteriores tem elevado potencial para novas pesquisas. O Quadro 2 apresenta os indicadores selecionados, de análise financeira, que foram utilizados, no presente estudo, com base no referencial levantado e nas delimitações da pesquisa.

Quadro 2 Indicadores

Cód.	Indicadores	Fórmula	Autores
X1	CDG sobre Ativo	CDG / AT	Apresentados neste trabalho com base nos estudos de Fleuriet et al. (2003)
X2	CDG sobre Receita Líquida	CDG / RL	
X3	NCG sobre Ativo	NCG / AT	
X4	NCG sobre Receita Líquida	NCG / RL	Brito, Assaf Neto e Corrar (2009), Minussi, Damacena e Ness Junior (2002)
X5	Saldo de Tesouraria sobre Ativo	T / AT	Brito, Assaf Neto e Corrar (2009), Horta (2010), Minussi, Damacena e Ness Junior (2002) e Sanvicente e Minardi (1998)
X6	Saldo de Tesouraria sobre Receita líquida	T / RL	
X7	Passivo Financeiro sobre Ativo Circulante	PF / AC	Eifert (2003)
X8	Tipo de Estrutura Financeira	TEF	Melo e Coutinho (2007)
X9	Termômetro de liquidez - TL	$T / (NCG)$	Horta (2010) e Vieira (2008)
X10	Saldo de Tesouraria	$T=AF-PF$	Melo e Coutinho (2007)
X11	Necessidade de Capital de Giro	$NCG = AO - PO$	Melo e Coutinho (2007)
X12	Capital de Giro	$CDG = PNC - ANC$	Melo e Coutinho (2007)
X13	Endividamento financeiro	$(PF + PNCF) / AT$	Brito, Assaf Neto e Corrar (2009)
X14	Capital de giro líquido	$(AC - PC) / AT$	Altman (1968), Brito, Assaf Neto e Corrar (2009) e Sanvicente e Minardi (1998)
X15	Retorno sobre o ativo	$LAJIR / AT$	Altman (1968), Altman, Baydia e Dias (1979), Brito, Assaf Neto e Corrar (2009)
X16	Patrimônio líquido dividido pelo exigível total	PL / ET	Altman, Baydia e Dias (1979), Horta (2010) e Soares e Rebouças (2015)
X17	Giro do ativo	RL / AT	Altman (1968), Brito, Assaf Neto e Corrar (2009) e Castro Junior (2003)
X18	Retorno sobre o Ativo	LL / AT	Castro Junior (2003), Eifert (2003), Horta (2010) e Soares e Rebouças (2015)
X19	Lucros retidos sobre ativo	$(LA + RL) / AT$	Brito, Assaf Neto e Corrar (2009)
X20	Rentabilidade do Patrimônio Líquido	LL / PL	Castro Junior (2003), Brito, Assaf Neto e Corrar (2009) e Kanitz (1978)
X21	Liquidez corrente	AC / PC	Castro Junior (2003), Horta (2010), Kanitz (1978), Matias (1978) e Silva (1983)
X22	Liquidez seca	$(AC - ESTOQUES) / PC$	Brito, Assaf Neto e Corrar (2009) e Horta (2010), Kanitz (1978)
X23	Liquidez geral	$(AC + RLP) / (PC + PNC)$	Kanitz (1978) e Soares e Rebouças (2015)

Fonte: Elaborado pelos autores

(continua)

Cód.	Indicadores	Fórmula	Autores
X24	Lucro operacional sobre despesas financeiras	LAJIR / DF	Brito, Assaf Neto e Corrar (2009) e Soares e Rebouças (2015)
X25	Margem Líquida	LL / RL	Brito, Assaf Neto e Corrar (2009), Castro Junior (2003), Elizabetsky (1976)
X26	Endividamento total	(PC + PNC) / AT	Brito, Assaf Neto e Corrar (2009), Horta (2010), Silva (1983)
X27	Indicador de estrutura	FORNECEDOR / AT	Matias (1978) e Silva (1983)
X28	Composição do Endividamento	PC / (PC + PNC)	Castro Junior (2003)
X29	Participação de Capitais de Terceiros	(PC + PNC) / PL	Castro Junior (2003)
X30	Imobilização do patrimônio líquido	ATIVO FIXO / PL	Brito, Assaf Neto e Corrar (2009), Castro Junior (2003) e Soares e Rebouças (2015)
X31	Patrimônio líquido sobre ativo	PL / AT	Brito, Assaf Neto e Corrar (2009) e Matias (1978)
X32	Estoques sobre ativo	ESTOQUES / AT	Brito, Assaf Neto e Corrar (2009), Elizabetsky (1976) e Silva (1983)
X33	Liquidez imediata	DISPONÍVEL / PC	Brito, Assaf Neto e Corrar (2009), Castro Junior (2003), Horta (2010)
X34	Indicador de estrutura	DISPONÍVEL / AT	Matias (1978), Silva (1983)
X35	Indicador de estrutura	DISPONÍVEL / ANC	Elizabetsky (1976), Silva (1983) e Soares e Rebouças (2015)

Legenda: AC = Ativo circulante, ACF = Ativo circulante financeiro, ACO = Ativo circulante operacional, AF = Ativo financeiro, AO = Ativo operacional, ANC = Ativo não circulante, AT = Ativo total, CDG = Capital de giro, DF = Despesas financeiras, EB = Estrutura de balanço (ordinal), ET = Exigível total, LA = Lucros acumulados, LAJIR = Lucro antes dos juros e imposto de renda, LB = Lucro Bruto, LL = Lucro líquido, NCG = Necessidade de capital de giro, PC = Passivo circulante, PF = Passivo financeiro, PL = Patrimônio líquido, PO = Passivo operacional, PNC = Passivo não circulante, PNCF = Passivo não circulante financeiro, RL = Reserva de Lucros, RLP = Realizável em longo prazo, T = Saldo em tesouraria, TL = Termômetro de liquidez, RL = Receita líquida.

Fonte: Elaborado pelos autores

Bellovary, Giacomino e Akers (2007) enfatizam que o número de indicadores utilizados nos modelos, também, é analisado por décadas, mostrando que a média tem variado ao longo do tempo, mas continua em torno de 10 indicadores no geral. Diante do exposto, com intuito de realizar modelos mais parcimoniosos, propõe-se utilizar de técnicas como o *stepwise*, que auxiliam na eliminação de indicadores menos significativos (com base em uma estatística F). Seguindo os passos de Soares e Rebouças (2015), outra possibilidade para se eliminar indicadores é por meio da análise fatorial, isto é, tendo a análise fatorial encontrado fatores (dimensões) com mais de uma variável, é possível retirar a variável com maior escore (poder discriminante) e utilizá-la, posteriormente, nas técnicas de classificação.

3.2 Técnicas para análise de crédito

Após a coleta dos dados amostrais (empresas solventes e insolventes), inicia-se à criação dos modelos e, com base nesta criação, realizam-se as suas respectivas análises e comparações. Para o presente estudo, são utilizados os modelos de análise discriminante (HAIR et al., 2009), regressão logística (HOSMER; LEMESHOW, 1989) e redes neurais (HAYKIN, 2001). Alguns exemplos práticos dessas metodologias podem ser vistos no trabalho de Castro Junior (2003), Iturriaga e Sanz (2015) e Soares e Rebouças (2015).

3.2.1 Avaliação dos modelos

Mesmo com resultados mais promissores, é preciso considerar as limitações de cada modelagem. Nas pesquisas empíricas que utilizam mais de uma técnica para avaliar o risco de

crédito é possível comparar a sua acurácia por meio de testes como a Curva ROC (*Receiver Operating Characteristics*) e o teste de Kolmogorov-Smirnov - KS. Segundo Crone e Finlay (2012) e Sicsú (2010), medidas de separação de grupos, tais como a área sob a curva ROC (AUC – *Area Under the Curve*) e a estatística KS, são amplamente utilizados para avaliar o desempenho dos modelos.

Sicsú (2010) e Sumihara Filho e Slegers (2010) salientam que alguns analistas costumam-se basear-se em alguns valores críticos de referência para o teste da AUROC e para o teste de Kolmogorov-Smirnov (KS), para determinar a eficácia dos modelos de *credit scoring*. Neste contexto, os autores observam que se pode utilizar um nível de precisão conforme o apresentado na Tabela 1.

Tabela 1 Poder discriminante: valores de referência da literatura

Hosmer e Lemeshow (<i>Applied Logistic Regression</i> , 1998, p. 162)		Sumihara Filho e Slegers (<i>Behaviour Scoring</i> , 2010, p. 44)	
<i>Área abaixo da curva ROC</i>	<i>Discriminação</i>	<i>Valores críticos para KS</i>	<i>Discriminação</i>
$\leq 0,5$	Não existe	$ks < 35$	Muito baixa
$> 0,5$ e $< 0,7$	Baixa	$35 \leq ks < 55$	Baixa
$\geq 0,7$ e $< 0,8$	Aceitável	$55 \leq ks < 65$	Aceitável
$\geq 0,8$ e $< 0,9$	Excelente	$65 \leq ks < 75$	Boa
$\geq 0,9$	Acima do comum	$ks \geq 75$	Excelente

Fonte: Adaptado de Sumihara Filho e Slegers (2010, p. 34-44)

Dentro da temática de risco de crédito e solvência, podem-se, ainda, citar alguns estudos recentes que utilizaram a curva ROC e o índice de Kolmogorov-Smirnov – KS, para avaliar a performance de modelos, entre eles destaca-se Ju e Sohn (2014) e Wosnitza e Leker (2014).

4 RESULTADOS E DISCUSSÃO

Como opção para definição dos indicadores, foi realizada a análise fatorial visando verificar entre os fatores encontrados quais indicadores seriam mais representativos do conjunto de dados. Por esse método obteve-se dois agrupamentos de indicadores, sendo. A primeira Fatorial com 7 indicadores (FATOR 01): Retorno sobre o Ativo (X18- LL / AT); Saldo de Tesouraria (X10- T = AF – PF); Indicador de estrutura (X34- DISPONÍVEL / AT); Necessidade de Capital de Giro sobre Receita líquida (X4- NCG / RL); Participação de Capitais de Terceiros (X29- (PC + PNC) / PL); Giro do ativo (X17- RL / AT) e Composição do Endividamento (X28- PC / (PC + PNC)).

A segunda Fatorial com 6 indicadores (FATORIAL 02): Necessidade de Capital de Giro sobre Receita Líquida (X4- NCG / RL); Saldo de Tesouraria (X10- T = AF – PF); Margem Líquida (X25- LL / RL); Imobilização do Patrimônio Líquido (X30- ATIVO FIXO / PL), Patrimônio Líquido sobre Ativo (X31- PL / AT), Indicador de Estrutura (X34- DISPONÍVEL / AT). Além dos indicadores com maior poder de explicação, encontrados pela análise fatorial, foram utilizados os escores das duas fatorais na tentativa de se encontrar a melhor combinação de indicadores para os modelos.

Neste sentido, têm-se, a princípio, sete possibilidades de combinações de indicadores para serem testadas: Resultado do método *Stepwise*, gerado por meio da Análise Discriminante, resultado do método *Forward*: LR gerado por meio da Regressão Logística, os indicadores mais representativos de cada fator, encontrados por meio da Fatorial 1, os indicadores mais representativos de cada fator encontrados por meio da Fatorial 2, o cálculo dos escores da Análise Fatorial 1, o cálculo dos escores da Análise Fatorial 2 e o processo humano de Tentativa e Erro que pode ser influenciado pelos resultados auferidos com as seis possibilidades utilizadas anteriormente.

4.1 Análise discriminante - AD

Para o modelo de análise discriminante, o melhor modelo foi encontrado por meio do processo de tentativa e erro. O parâmetros relevantes para a modelagem foram: indicadores significantes (sig. < 0,05); Correlação canônica $(0,814)^2$ explicando 66% da discriminação entre os grupos; nenhuma correlação mais acentuada de Multicolinearidade (valores de tolerância > 0,1 e valores de VIF < 10).

A análise dos Coeficientes da função de classificação permite conhecer um pouco mais sobre qual a importância de cada indicador dentro da Função Discriminante, os indicadores que apresentaram valores negativos para o coeficiente da função discriminante foram: Capital de Giro sobre Ativo (X1- CDG / AT) e Tipos de Estrutura Financeira (X8- TEF = Tipos de Estrutura Financeira), contribuirão para que a empresa seja classificada como insolvente. Por outro lado, os indicadores que apresentarem valores positivos para o coeficiente da função discriminante foram: Necessidade de Capital de Giro sobre Ativo (X3- NCG / AT); Termômetro de Liquidez (X9- $TL = T / (|NCG|)$); Rentabilidade do Patrimônio Líquido (X20- LL / PL); Margem Líquida (X25- LL / RL), Patrimônio líquido sobre ativo (X31- PL / AT) colaborarão para que a empresa seja classificada como solvente.

De posse dos coeficientes da função discriminante canônica não padronizado, pode-se elaborar a função para a Análise Discriminante, ou seja, o *Credit Scoring* gerado pode ser representado pela Equação 1.

$$Z = 1,649 - 0,180 \cdot \left(\frac{CDG}{AT}\right) + 1,207 \cdot \left(\frac{NCG}{AT}\right) - 0,449 \cdot (Tipo\ de\ Estrutura) + 0,046 \cdot \left(\frac{T}{|NCG|}\right) + 1,036 \cdot \left(\frac{LL}{PL}\right) + 1,627 \cdot \left(\frac{LL}{VL}\right) + (0,693 \cdot \left(\frac{PL}{AT}\right)) \quad (1)$$

Elaborada a função para a Análise Discriminante é possível calcular o ponto de corte a partir dos centroides de cada grupo. Os centroides são as médias obtidas com a distribuição individual dos grupos. A média ponderada entre os centroides de cada uma das distribuições se constituirá no ponto de corte da função discriminante. Neste sentido, o resultado obtido para o ponto de corte ótimo é $-0,4978$, este valor servirá para classificar as empresas por meio do seu escore discriminante. Ou seja, as empresas que se encontrarem abaixo do ponto de corte ($-0,4978$) serão classificadas como pertencentes ao grupo "1" (insolventes) e as empresas que possuem escore discriminante acima do ponto do corte irão ser classificadas como parte do grupo "0" (solventes).

4.2 Regressão logística - RL

Para o cálculo da Regressão Logística, o método *Forward*: LR apresentou os melhores resultados. A validação dos resultados foi obtido pelos parâmetros: Cox & Snell 59%; Nagelkerke 79%; Log-verossimilhança -2 (-2VL) com 64,5% de explicação dos motivos que levam uma empresa a se tornar insolvência; Hosmer e Lemeshow Sig. > 0,05 e Multicolinearidade (valores de tolerância > 0,1 e valores de VIF < 10).

Os resultados encontrados para os indicadores no modelo apresentam aspectos positivos ao tentar estimar a probabilidade a partir de cada coeficiente em particular, como se observa na Tabela 2. A Estatística Wald apresenta um coeficiente Wald maior que zero (wald > 0), para cada fator, exercendo efeito discriminatório sobre a probabilidade de a empresa ser insolvente ou solvente, os indicadores, também, são significativos a 0,05 (Sig. < 0,05). Os coeficientes dos fatores independentes, que se encontram na coluna Exp(B) (25,34), estão dentro do limite estabelecido pelas colunas Inferior e Superior (3,12 a 205,38 respectivamente), assim como todos os outros fatores utilizados.

Tabela 1 Indicadores na equação

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
X3NCGAT	-10,747	3,480	9,537	1	0,002	0,000	0,000	0,020
X16PLET	-2,800	0,731	14,670	1	0,000	0,061	0,015	0,255
X25LLRL	-7,329	1,887	15,089	1	0,000	0,001	0,000	0,026
X28PCPCPNC	3,232	1,068	9,168	1	0,002	25,343	3,127	205,379

a. Indicadores inseridos na etapa 1: X3NCGAT, X16PLET, X25LLRL, X28PCPCPNC

Fonte: Dados da Pesquisa

O modelo de Regressão Logística capaz de classificar as empresas em insolvente e solvente utiliza-se dos coeficientes da coluna B, apresentados na Tabela 2. A equação da Regressão Logística, ou *Credit Scoring* gerado, pode ser representada conforme Equação 2.

$$P(Y=1) = \frac{1}{1 + e^{-(-10,747 \cdot (\frac{NCG}{AT}) - 2,800 \cdot (\frac{PL}{ET}) - 7,329 \cdot (\frac{LL}{VL}) + 3,232 \cdot (\frac{PC}{PC+PNC}))}} \quad (2)$$

Para a equação da regressão logística, observa-se que $P(Y=1)$ apresenta saída do tipo dicotômica, assumindo apenas dois valores possíveis '0' ou '1'. Neste sentido, quanto mais próximo de 1 for o resultado mais a empresa será considerada insolvente, do contrário, quanto menor for o resultado, aproximando-se de 0, mais a empresa será tratada como solvente. Como o resultado oscila entre '0' e '1', o ponto de corte para a regressão logística é 0,5. Pode -se, ainda, multiplicar este resultado por 100 e interpretá-lo como probabilidade (em %) de uma empresa ser insolvente.

4.3 Redes neurais artificiais - RNA

Para o modelo com redes neurais artificiais, o melhor modelo obtido foi por meio do processo de tentativa e erro. Para modelagem utilizou-se uma rede *perceptron* de multicamadas (MLP) com arquitetura *feedforward* e algoritmo back-propagation que atingiu critério do erro de treinamento em 0,0001. As informações da rede são apresentadas na Tabela 3. As covariáveis (número de oito unidades) representam os indicadores selecionados. Para a arquitetura, foi definida uma camada oculta contendo sete neurônios (unidades). A função de ativação, para camada oculta, foi a função *Tangente hiperbólica* e, para a camada de saída, utilizou-se a função de ativação *Softmax*. As funções de ativação foram definidas por apresentar melhor desempenho entre as funções testadas.

Tabela 2 Informações de rede

Camada de entrada	Covariáveis	1	Modelo	X4) NCG / RL
		2		X9) TL = T / (NCG)
		3		X16) PL / ET
		4		X17) RL / AT
		5		X20) LL / PL
		6		X25) LL / RL
		7		X28) PC / (PC + PNC)
		8		X31) PL / AT
Número de unidades ^a		8		
Método de reescalonamento para covariáveis		Padronizado		
Camadas ocultas	Número de camadas ocultas		1	
	Número de Unidades na Camada Oculta 1		7	
	Função de ativação		Tangente hiperbólica	
Camada de saída	Variáveis dependentes	1	Situação 0 1	
	Número de unidades		2	
	Função de ativação		Softmax	
	Função de erro		Entropia cruzada	

Fonte: Dados da pesquisa

Os indicadores mais importantes, para o modelo de redes neurais, por ordem de importância, foram: Termômetro de liquidez (X09- $TL = T / (|NCG|)$); Necessidade de Capital de Giro sobre Receita líquida (X03- NCG / RL); Patrimônio líquido sobre ativo (X31- PL / AT); Rentabilidade do Patrimônio Líquido (X20- LL / PL); Margem Líquida (X25 - LL / RL); Giro do ativo (X17- RL / AT); Patrimônio Líquido sobre Exigível Total (X16- PL / ET) e Composição do Endividamento (X28- $PC / (PC + PNC)$).

Contudo, a importância de um indicador independente é calculada com a variação sofrida pela rede em seu valor previsto (saída), para diferentes valores de entrada. Não se pode afirmar é que a ‘direção’ que os indicadores contribuirão para probabilidade da empresa se solvente ou insolvente, isto é, não se pode afirmar que um maior valor para o indicador X, indica uma maior probabilidade da empresa se tornar insolvente.

4.4 Comparação dos modelos

Para comparação dos modelos, a Tabela 4 apresenta, primeiramente, o nível de acurácia alcançado em cada agrupamento de indicadores para os três modelos testados. Observa-se que o nível geral de acertos para todas as combinações de indicadores foi acima de 80%, de precisão, o que demonstra que as sete opções utilizadas, para agrupar os indicadores, foram bastante relevantes e são passíveis de serem utilizadas para auxiliar o especialista ou estudioso do campo do risco de crédito, no que tange à escolha de quais indicadores utilizar para as análises.

A coluna ‘Tentativa e erro’, na Tabela 4, foi elaborada com intuito de se buscar melhores parâmetros para cada modelagem, ou seja, modelos mais parcimoniosos, com maior acurácia nas previsões e com baixos níveis de erro tipo 1 (Erro tipo I: classificar empresas insolventes como solventes e o ato da concessão de empréstimo ocorrer em perdas financeiras; erro tipo II: classificar empresas solventes como insolventes e não conceder o crédito, deixando, assim, de auferir lucros). Esse processo de tentativa e erro é o que normalmente ocorre, quando o especialista usa de seu conhecimento e intuição, para alcançar resultados mais precisos, tendo em vista seu conhecimento específico de determinado assunto.

Tabela 3 Comparação dos modelos com indicadores diferentes por modelo

Modelo	Stepwise	Forward: LR	Indicadores da F1	Indicadores da F2	Fatorial 1	Fatorial 2	Tentativa e erro
	% de acerto	% de acerto	% de acerto	% de acerto	% de acerto	% de acerto	% de acerto
Análise discriminante ¹	89,0	86,2	88,0	86,1	89,8	86,0	90,9
Regressão logística	90,1	90,9	86,0	87,6	89,3	86,0	-
Redes neurais artificiais ¹	88,6	89,2	86,4	89,8	92,4	87,9	97,8

Obs.: ¹ Para os modelos, foi utilizada a média aritmética simples entre: o resultado da amostra de treinamento e o resultado da amostra de validação

Fonte: Elaborada pelos autores

Os agrupamentos de indicadores mais significativos foram a técnica *Forward*, utilizada para regressão logística, com 90,9% de acurácia, também, o processo de tentativa e erro, realizado para a análise discriminante, com 90,9% de precisão e, para as redes neurais, com 97,8% de acerto (TABELA 4). Cabe observar que todos os agrupamentos realizados contribuíram para se definir os indicadores por meio do processo de tentativa e erro.

A acurácia dos modelos selecionados, quanto ao nível de Precisão Tipo I e Precisão Tipo 2, pode ser observada na Tabela 5. A Precisão Tipo I é o oposto do Erro Tipo I, isto é, quanto maior o nível de Precisão Tipo I, menor o nível de Erro Tipo I, proporcionalmente – o mesmo para a Precisão Tipo II. Na Tabela 5, pode-se ver que a análise discriminante e a

regressão logística apresentaram os mesmos níveis de acerto, 90,9%. Contudo, ao se avaliar os níveis de precisão tipo I e tipo II, fica evidente que o modelo de regressão logística apresenta melhor classificação quanto à precisão tipo I, ou seja, o percentual de classificação de empresas insolventes como solventes é maior para o modelo de análise discriminante do que para o modelo de regressão logística, o que acarreta perda do capital investido.

Tabela 5 Acurácias obtidas pelos três modelos finais (aos níveis de erro tipo I e tipo II)

Modelo	Precisão Tipo I	Precisão Tipo II	Precisão Geral
Análise discriminante ¹	82,4%	97,1%	90,9%
Regressão logística	90,2%	91,4%	90,9%
Redes neurais artificiais ¹	98,0%	98,6%	97,8%

Obs.: ¹ Para os modelos, foi utilizada a média aritmética simples entre: o resultado da amostra de treinamento e o resultado da amostra de validação

Fonte: Elaborada pelos autores

Os modelos, também, podem ser avaliados por meio da estatística Kolmogorov-Smirnov (KS), que busca verificar o comportamento dos escores das duas amostras de cada modelo. Segundo Selau (2008, p. 96), “espera-se comprovar uma homogeneidade da distribuição dos resultados dos escores nas duas amostras, necessária para utilização do modelo como previsão do risco de crédito a novos clientes”. De posse dos valores, para as distribuições dos modelos gerados, pode facilitar a compreensão do teste, ao gerar um gráfico com a visualização do grau de separação entre as empresas solventes e as empresas insolventes por meio dos escores dos modelos (Figura 1).

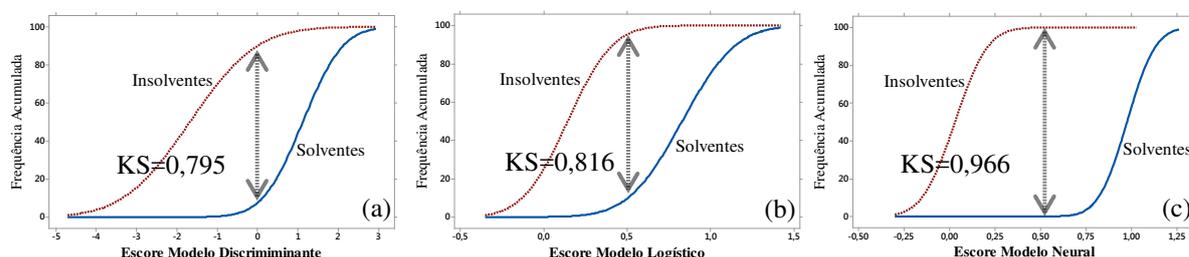


Figura 1 Distribuição das frequências acumuladas para KS para os modelos construídos (a = Análise Discriminante; b = Regressão Logística e c = Redes Neurais)

Fonte: Elaborada pelos autores (Gráficos do *software* Minitab)

O valor da estatística KS, ainda, pode ser comparado a um nível de significância de $\alpha = 0,01$, que pode ser obtido pelo cálculo do ponto crítico: $KS_{1\%} = \frac{1,63 \sqrt{n_1 + n_2}}{\sqrt{n_1 * n_2}} = \frac{1,63 \sqrt{70 + 51}}{\sqrt{70 * 51}} = 0,1254$. Neste sentido, observa-se que todos os valores máximos encontrados para KS são maiores que o valor D-Crítico de 0,1254. Dado que o valor observado da estatística de teste D-Máximo > D-Crítico, rejeita-se, a hipótese nula de que as distribuições de empresas solventes e empresas insolventes é a mesma. Há, portanto, evidência estatística, ao nível de significância de 0,01, de que os valores da análise se distribuem de forma diferente para o grupo das empresas solventes e insolventes da amostra.

A Tabela 1 apresenta os valores para a estatística KS, segundo Sumihara Filho e Slegers (2010), valores críticos de KS acima de 0,75 são considerados ‘excelentes’ e quanto mais próximo de 1 os valores para o teste, maior será a distancia entre as distribuições acumuladas dos scores dos modelos gerados. Neste contexto, pode-se ver que todos os modelos apresentaram valores superiores a 0,75, contudo, fica evidente a superioridade da Rede Neural, ao apresentar valores de 0,966, em segundo, segue a Regressão Logística com valores de 0,816 e, por último, a Análise Discriminante como valores de 0,795 (TABELA 6).

Tabela 4 Estatísticas de teste KS para os modelos construídos

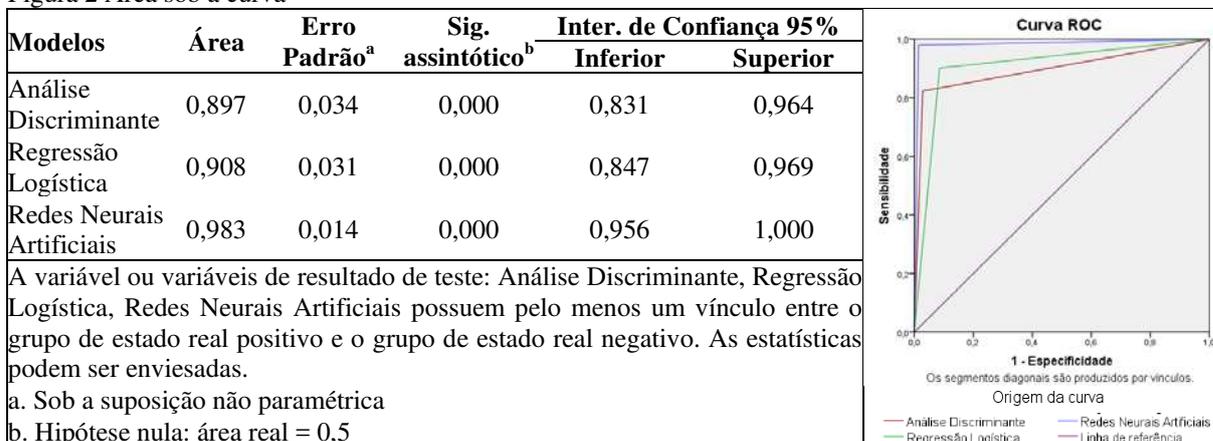
		Análise Discriminante	/Regressão Logística	/Redes Neurais Artificiais
Diferenças Mais Extremas	Absoluto	0,795	0,816	0,966
	Positivo	0,795	0,816	0,966
	Negativo	0,000	0,000	0,000
Z Kolmogorov-Smirnov		4,318	4,434	5,248
Significância Sig. (2 extremidades)		0,000	0,000	0,000

Fonte: Elaborada pelos autores

Além do teste KS, autores como Baensens et al. (2003) e Sicsú (2010) afirmam que, para realizar avaliação do poder discriminador dos modelos de classificação, os analistas costumam utilizar, também, os valores da curva ROC (*Receiver Operating Characteristic Curves*), a qual tem encontrado bons resultados no campo de risco de crédito. Na Figura 2, podem-se ver as curvas ROC, obtidas para os três melhores modelos gerados. Observa-se que a análise discriminante obteve a menor área sob a curva. Já a regressão logística apresentou um resultado melhor do que a análise discriminante. Contudo, o melhor desempenho foi apresentado pelas redes neurais, ocupando quase toda a área da curva ROC.

Segundo Sumihara Filho e Slegers (2010), como demonstrado na Tabela 1, pode-se classificar os modelos com base nos valores obtidos para a área da curva ROC como: $roc \leq 0,5$ não existe; $0,5 < roc < 0,7$ = baixa; $0,7 \leq roc < 0,8$ = aceitável; $0,8 \leq roc < 0,9$ = excelente e $roc \geq 0,9$ = acima do comum. Neste sentido, como pode ser visto na Figura 2, o modelo de análise discriminante foi classificado sob a área da curva ROC como 0,897, ou seja, considerados como excelente. Já os modelos de regressão logística e redes neurais foram classificados com 0,908 e 0,983, respectivamente. Ambos os modelos estão acima do esperado em questão de classificação, apresentando resultados acima de excelente.

Figura 2 Área sob a curva



Fonte: Elaborada pelos autores

A Tabela 7 apresenta uma síntese com as medidas de desempenho (Nível de Acurácia, Erro Tipo I, Teste KS e Curva ROC) utilizadas para comparar os modelos finais encontrados. Neste resumo, pode-se observar que as Redes Neurais Artificiais apresentaram melhores resultados em todos os testes realizados, ou seja, alto nível de acurácia, baixo nível de erro Tipo I e altos valores para o teste KS e para a curva ROC.

Tabela 7 Medidas de desempenho para os modelos gerados

Modelos	Nível de Acurácia	Erro Tipo I	Teste KS	Curva ROC
Análise Discriminante	90,9	17,6%	0,795	0,897
Regressão Logística	90,9	9,8%	0,816	0,908
Redes Neurais Artificiais	97,8	2%	0,966	0,983

Fonte: Elaborada pelos autores

4.5 Análise dos Indicadores

A Tabela 8 apresenta os indicadores que foram mais representativos para todos os agrupamentos encontrados (todos os agrupamentos de indicadores utilizados alcançaram precisão de acerto acima de 86%). Os três modelos finais que apresentaram melhores resultados foram *Forward*: LR para regressão logística, Tentativa e erro, para análise discriminante e Tentativa e erro, para as redes neurais artificiais (com os indicadores ‘X’ destacados em negrito na Tabela 8, também se destacou os indicadores que apareceram três ou mais vezes).

Tabela 8 Indicadores selecionados

Indicadores	<i>Forward</i> : LR	<i>Stepwise</i>	Tentativa e erro, AD	Tentativa e erro, RNA	FATORIAL 01	FATORIAL 02	Total de repetições
X1) CDG / AT			X				1
X3) NCG / AT	X		X				2
X4) NCG / RL		X		X	X	X	4
X8) TEF = Estrutura		X	X				2
X9) TL = T / (NCG)		X	X	X			3
X10) T = AF – PF					X	X	2
X11) NCG = AO – PO		X					1
X16) PL / ET	X			X			2
X17) RL / AT				X	X		2
X18) LL / AT		X			X		2
X20) LL / PL		X	X	X			3
X25) LL / RL	X		X	X		X	4
X28) PC / (PC + PNC)	X			X	X		3
X29) (PC + PNC) / PL					X		1
X30) ATIVO FIXO / PL						X	1
X31) PL / AT			X	X		X	3
X34) DISPONÍVEL / AT					X	X	2

Fonte: Elaborada pelos autores

Neste sentido, pode-se dizer que os indicadores selecionados, para os três modelos finais, que apresentaram baixos valores para as empresas insolventes e altos valores para as empresas solventes foram: Necessidade de Capital de Giro sobre Ativo (X3); Termômetro de Liquidez (X9); Rentabilidade do Patrimônio Líquido (X20); Margem Líquida (X25) e Patrimônio líquido sobre ativo (X31). Por outro lado, apenas dois indicadores apresentaram valores maiores para as empresas insolventes: Capital de Giro sobre Ativo (X1) e Tipo de Estrutura Financeira (X8). No total, dos 35 indicadores selecionados inicialmente, apenas 17 participaram dos agrupamentos para análise.

5 CONCLUSÃO

Todos os pressupostos para a modelagem foram atendidos. Ao todo, para os três modelos e os sete agrupamentos de indicadores propostos, foram gerados 21 modelos (além dos diversos testes realizados por meio de tentativa e erro), os quais apresentaram acurácia acima de 85% em suas classificações, o que demonstra que, em parte, os agrupamentos propostos obtiveram ‘bons’ resultados.

Entretanto, no decorrer dos resultados, apenas o melhor modelo para cada técnica foi apresentado. Ao se comparar as técnicas utilizadas para prever a insolvência empresarial, pode-se perceber a superioridade das redes neurais artificiais sobre a regressão logística e desta sobre a análise discriminante. Estes resultados foram confirmados pelo nível de acertos apresentados pelas técnicas: análise discriminante 90,9%; regressão logística 90,9%; e as redes neurais artificiais 97,8%.

Outro fato que merece destaque é o nível de acurácia obtido pelos três modelos finais no que se refere à Precisão Tipo I (menor nível de erro tipo I), em que a análise discriminante obteve 82,4%, a regressão logística 90,2%, e as redes neurais artificiais alcançaram 98,0%. A estatística de Kolmogorov-Smirnov (KS), também, apresentou valores excelentes ($ks \geq 75$) para Análise Discriminante (0,795), Regressão Logística (0,816) e Redes Neurais Artificiais (0,966). Isso foi atestado, por meio da curva ROC, ao apresentar valores excelentes para a análise discriminante (0,897), e valores acima do comum para a regressão logística (0,908) e para as redes neurais artificiais (0,983).

Uma das limitações da pesquisa constituiu-se no fato de não ter sido possível realizar os modelos por setores específicos, o que poderia contribuir tanto para a precisão dos modelos quanto para uma melhor compreensão dos fatores que determinam a insolvência das companhias de capital aberto no Brasil. Como recomendações para pesquisas futuras indica-se, a possibilidade de se desenvolver modelos híbridos, utilizando técnicas de Inteligência Computacional que possibilitem uma melhor precisão para a modelagem do risco de crédito.

Acredita-se que, por meio do presente estudo, pode-se elucidar algumas das características que possuem as empresas insolventes para a presente amostra. Tais contribuições são fundamentais para as pesquisas de risco de crédito e contribuem para o desenvolvimento da temática ao se comparar técnicas de análise (AD, RL e RNA).

REFERÊNCIAS

- ALTMAN, E. I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. **The Journal of Finance**, New York, 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. **Revista de Administração de Empresas**, São Paulo, v. 19, n. 1, p. 17-28, 1979.
- ASSAF NETO, A. **Finanças corporativas e valor**. 5. ed. São Paulo: Atlas, 2010. 726 p.
- ASSAF NETO, A.; SILVA, C. A. T. **Administração do capital de giro**. 4. ed. São Paulo: Atlas, 2012. 269 p.
- BAESENS, B. et al. Benchmarking state-of-the-art classification algorithms for credit scoring. **Journal of the Operational Research Society**, v. 54, n. 6, p. 627-635, 2003.
- BEAVER, W. H. Financial ratios as predictors of failure. **Journal of Accounting Research**, Chicago, v. 4, p. 71-111, 1966.
- BELLOVARY, J. L.; GIACOMINO, D. E.; AKERS, M. D. A review of bankruptcy prediction studies: 1930 to present. **Journal of Financial Education**, v. 33, p. 1-42, 2007.
- BRASIL. Congresso Nacional. **Lei nº 11.101**, de 9 de fevereiro de 2005. Brasília, 2005. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2005/Lei/L11101.htm#art200>. Acesso em: 22 dez. 2015.
- BRASIL. Presidência da República. **Decreto-Lei nº 7.661**, de 21 de junho de 1945. Lei de Falências, 1945. Disponível em: <<http://www.planalto.gov.br>>. Acesso em: 22 dez. 2015.
- BRITO, G. A. S.; ASSAF NETO, A.; CORRAR, L. J. Sistema de classificação de risco de crédito: uma aplicação a companhias abertas no Brasil. **Revista Contabilidade & Finanças**, São Paulo, v. 20, n. 51, p. 28-43, 2009.
- CASTRO JUNIOR, F. H. F. **Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais**. 2003. 169 p. Dissertação (Mestrado em Administração) - Universidade de São Paulo, São Paulo, 2003.
- CHUANG, C. L.; HUANG, S. T. A hybrid neural network approach for credit scoring. **Expert Systems**, New York, v. 28, n. 2, p. 185-196, 2011.
- CRONE, S. F.; FINLAY, S. Instance sampling in credit scoring: an empirical study of sample size and balancing. **International Journal of Forecasting**, v. 28, n. 1, p. 224-238, 2012.
- DAMODARAN, A. **Investment Valuation: Tools and Techniques for Determining the Value of Any Asset**. 3. ed. New York: John Wiley & Sons: 2011. 992 p.

DURAND, D. Risk elements in consumer installment lending. **Studies in Consumer Installment Financing**, New York, v. 8, p. 1-101, 1941.

EIFERT, D. S. **Análise quantitativa na concessão de crédito versus inadimplência**: um estudo empírico. 2003. 81 p. Dissertação (Mestrado em Administração) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

ELIZABETSKY, R. **Um modelo matemático para decisão no banco comercial**. 1976. 190 p. Dissertação (Mestrado em Engenharia de Produção) - Escola Politécnica, Universidade de São Paulo, São Paulo, 1976.

FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, London, v. 7, n. 2, p. 179-188, 1936.

FLEURIET, M. et al. **O Modelo Fleuriet**: a dinâmica financeira das empresas brasileiras: um método de análise, orçamento e planejamento financeiro. 3. ed. Rio de Janeiro: Campus, 2003. 169 p.

GARCIA, F.; GIMENEZ, V.; GUIJARRO, F. Credit risk management: a multicriteria approach to assess creditworthiness. **Mathematical and Computer Modelling**, Oxford, v. 57, n. 7/8, p. 2009-2015, Apr. 2013.

HAIR, J. F. et al. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

HARRIS, T. Quantitative credit risk assessment using support vector machines: broad versus Narrow default definitions. **Expert Systems With Applications**, New York, v. 40, n. 11, p. 4404-4413, 2013.

HAYKIN, S. **Redes neurais**: princípios e prática. 2. ed. Porto Alegre: Bookman, 2001. 900 p.

HORTA, R. A. M. **Uma metodologia de mineração de dados para a previsão de insolvência de empresas brasileiras de capital aberto**. 2010. 151 p. Tese (Doutorado em Engenharia Civil) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

HOSMER JUNIOR, D. W.; LEMESHOW, S. **Applied logistic regression**. New York: J. Wiley, 2004. 383 p.

ITURRIAGA, F. J. L.; SANZ, I. P. Bankruptcy visualization and prediction using neural networks: a study of US commercial banks. **Expert Systems with Applications**, New York, v. 42, n. 6, p. 2857-2869, 2015.

JU, Y. H.; SOHN, S. Y. Updating a credit-scoring model based on new attributes without realization of actual data. **European Journal of Operational Research**, Amsterdam, v. 234, n. 1, p. 119-126, 2014.

KANITZ, S. C. Como prever falências de empresas. **Revista Negócios em Exame**, São Paulo, p. 95-102, dez. 1974.

KANITZ, S. C. **Como prever falências**. São Paulo: McGraw-Hill do Brasil, 1978. 177 p.

KIM, S. Y. Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis. **Service Industries Journal**, London, v. 31, n. 3, p. 441-468, 2011.

KIMURA, H. et al. **Value-at-Risk**: como entender e calcular o risco pelo VaR: uma contribuição para a gestão no Brasil. Ribeirão Preto: Inside Books, 2008. 223 p.

KNIGHT, F. H. **Risk, uncertainty and profit**. 5. ed. Washington: Beard Books, 2002. 173 p.

KOU, G.; WU, W. An analytic hierarchy model for classification algorithms selection in credit risk analysis. **Mathematical Problems in Engineering**, Cairo, v. 2014, p. 1-7, 2014.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 6. ed. São Paulo: Atlas, 2005. 315 p.

MARTINS, G. A.; THEÓPHILO, C. R. **Metodologia da investigação científica para ciências sociais aplicadas**. São Paulo: Atlas, 2009. 264 p.

MARTINS, M. S. **A previsão de insolvência pelo modelo de Cox**: uma contribuição para a análise de companhias abertas brasileiras. 2003. 102 p. Dissertação (Mestrado em Administração) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

MATIAS, A. B. **Contribuição às técnicas de análise financeira: um modelo de concessão de crédito.** 1978. 101 p. Monografia (Graduação em Administração) - Universidade de São Paulo, São Paulo, 1978.

MELO, A. C.; COUTINHO, E. S. O modelo Fleuriet como indicador conjunto de solvência e rentabilidade. In: ENCONTRO DA ANPAD, 31., 2007, Rio de Janeiro. **Anais...** Rio de Janeiro, 2007. 1 CD-ROM.

MINUSSI, J. A.; DAMACENA, C.; NESS JUNIOR, W. L. Um modelo de previsão de solvência utilizando regressão logística. **Revista de Administração Contemporânea**, Curitiba, v. 6, n. 3, p. 109-128, 2002.

ODOM, M. D.; SHARDA, R. A neural network model for bankruptcy prediction. In: IJCNN INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 1990, San Diego. **Proceedings...** San Diego, 1990. p. 163-168.

OHLSON, J. Financial ratios and the probabilistic prediction of bankruptcy. **Journal of Accounting Research**, Chicago, v. 18, n. 1, p. 109-131, 1980.

ORESKI, S.; ORESKI, G. Genetic algorithm-based heuristic for feature selection in credit risk assessment. **Expert Systems with Applications**, v. 41, n. 4, p. 2052-2064, 2014.

PEREIRA, J. M.; DOMÍNGUEZ, M. Á. C.; OCEJO, J. L. S. Modelos de previsão do fracasso empresarial: aspectos a considerar. **Tékhné-Revista de Estudos Politécnicos**, Botucatu, n. 7, p. 111-148, 2007.

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

SABATO, G. Modelos de Scoring de risco de crédito. **Revista Tecnologia de Crédito**, São Paulo, v. 1, n. 68, p. 29-47, 2009.

SANVICENTE, A. Z.; MINARDI, A. M. A. F. **Identificação de indicadores contábeis significativos para a previsão de concordata de empresas.** São Paulo: Instituto Brasileiro de Mercado de Capitais, 1998. 12 p. Working Paper.

SCHRICKEL, W. K. **Análise de crédito: concessão e gerencia de empréstimos.** 5. ed. São Paulo: Atlas, 2000. 353 p.

SELAU, L. P. R. **Construção de modelos de previsão de risco de crédito.** 2008. 129 p. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.

SICSÚ, A. L. **Credit scoring: desenvolvimento, implantação, acompanhamento.** São Paulo: Blucher, 2010. 180 p.

SILVA, J. P. **Administração de crédito e previsão de insolvência.** São Paulo: Atlas, 1983.

SILVA, J. P. **Gestão e análise de risco de crédito.** 6. ed. São Paulo: Atlas, 2008. 421 p.

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**, Rio de Janeiro, v. 18, n. 3, p. 40-61, 2015.

SUMIHARA FILHO, H.; SLEEGERS, L. C. Valores de referência para os principais indicadores de acurácia dos modelos de escoragem. **Serasa Experian - Tecnologia de Crédito**, São Paulo, n. 73, p. 31-45, ago. 2010.

VERGARA, S. C. **Projetos e relatórios de pesquisa em administração.** 9. ed. São Paulo: Atlas, 2008. 94 p.

VIEIRA, M. V. **Administração estratégica do capital de giro.** 2. ed. SP: Atlas, 2008. 326 p.

WOSNITZA, J. H.; LEKER, J. Why credit risk markets are predestined for exhibiting log-periodic power law structures. **Physica A: Statistical Mechanics and its Applications**, London, v. 393, p. 427-449, Jan. 2014.

YU, L. Credit risk evaluation with a least squares fuzzy support vector machines classifier. **Discrete Dynamics in Nature and Society**, New York, v. 2014, p. 1-9, 2014.