

Uma revisão sistemática da literatura sobre Credit Scoring**A systematic review of literature on Credit Scoring**

DOI:10.34117/bjdv6n3-009

Recebimento dos originais: 30/02/2020

Aceitação para publicação:03/03 /2020

Flávio Fuhr

Mestre em Engenharia de Produção e Sistemas pela Universidade Tecnológica Federal do Paraná, Câmpus Pato Branco.

Universidade Tecnológica Federal do Paraná, Câmpus Pato Branco Country: Brasil
Endereço Institucional: Av. Bento Munhoz da Rocha Neto s/nº - PRT-280, Trevo da Codapar, Palmas - PR, 85555-000.

E-mail: flavio.fuhr@ifpr.edu.br

José Donizetti de Lima

Doutor em Engenharia de Produção pela Universidade Federal do Rio Grande do Sul
Universidade Tecnológica Federal do Paraná, Câmpus Pato Branco,

Endereço Institucional: Via do Conhecimento, s/n - KM 01 - Fraron, Pato Branco - PR,
85503-390.

E-mail: donizetti@utfpr.edu.br

Fernando José Avancini Schenatto QJosé Donizetti

Doutor em Engenharia de Produção pela Universidade Federal de Santa Catarina.

Universidade Tecnológica Federal do Paraná, Câmpus Pato Branco ,Brasil
Endereço Institucional: Via do Conhecimento, s/n - KM 01 - Fraron, Pato Branco - PR,
85503-390.

E-mail: Schenatto@utfpr.edu.br

RESUMO

A pontuação de crédito, “*credit scoring*”, transformou-se em uma técnica de vital importância junto as instituições fomentadoras de crédito na busca por avaliar e gerenciar o risco de crédito. O objetivo deste trabalho foi de realizar uma revisão sistemática da literatura sobre as técnicas utilizadas para esse fim nos últimos 09 anos, diretamente em artigos, principalmente os voltados a área de finanças e falências. Para a seleção do material utilizado junto aos bancos de pesquisa, a metodologia adotada foi o *Proknow- C (Knowledge Development Process – Constructivist*. Os resultados mostram que a técnica de Regressão Logística (RL) e Análise Discriminante (AD) são bastante utilizadas, tanto na predição direta como na comparação com outros resultados, havendo crescente interesse pelas técnicas computacionais como Redes Neurais Artificiais (RNA) e *Support Vector Machines (SVM)*.

Palavras-Chave: Análise de Crédito, *Credit Scoring*, Técnicas Estatísticas e Computacionais.

ABSTRACT

Credit scoring has become a vitally important technique among credit-promoting institutions in their quest to measure and manage credit risk. The objective of this work was to perform a systematic review of the literature on the techniques used for this purpose in the last 09 years, directly in articles, especially those focused on finance and bankruptcy. For the selection of material used by the research banks, the methodology used was Proknow-C (Knowledge Development Process - Constructivist). The results show that the Logistic Regression (RL) and Discriminant Analysis (AD) techniques are widely used, both in direct prediction and in comparison with other results, with increasing interest in computational techniques such as Artificial Neural Networks (RNA) and Support Vector Machines (SVM)

Key-words: Credit Analysis, Credit Scoring, Statistical and Computational Techniques.

1 INTRODUÇÃO

O crédito tem papel relevante na economia, sua ampliação no mercado do Brasil, é um dos responsáveis para o crescimento da economia, conforme destaca o Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE, 2014).

O outro lado do crédito é destacado pelo risco proporcionado pelo empréstimo. A preocupação com o fornecimento de crédito e sua importância para sustentabilidade do sistema comercial já era presente em 1938, quando a instituição “*National Bureau of Economic Research*” promoveu estudos em instituições de fomento de empréstimos nos quais buscou-se criar um histórico do crédito de bons e maus pagadores utilizando medidas estatísticas, a fim de identificar o risco do crédito (DURAND, 1941).

Esses modelos ganharam maior relevância em épocas de crises econômicas, como as ocorridas próxima e pós década de 1990 nos Estados Unidos, México, Brasil, Rússia, entre outros países. Devido as relações financeiras, essas crises afetaram outros países relacionados, prejudicando o crescimento econômico e causando efeitos como fechamento de empresas e postos de trabalho (LOUZADA *et al.*, 2016).

Na última década, mais especificamente no ano de 2008, o Brasil foi atingido novamente por mais uma crise internacional, chamada de crise do *subprime*, deflagrada nos Estados Unidos, a qual aprofundou-se, havendo fuga de capitais e desvalorização das moedas das economias periféricas. O agravamento da crise financeira internacional, levou ao extremo a aversão aos riscos pelas instituições financeiras (PRATES e FARHI, 2009).

As instituições financeiras, visando melhorar a confiança no sistema financeiro, na tentativa de mitigar crises econômico-financeiras, buscam emitir práticas e técnicas que visem dar segurança ao mercado. Nesse contexto, em 1988, na Suíça foi lançado um acordo, a serem observados pelas organizações creditícias, conhecido como: *International Convergence of Capital Measurement and Capital Standards*, mas difundiu-se como Basileia I, tendo seu 2º acordo “Basileia II”, lavrado em 2004 e mais recentemente o 3º acordo, “Basileia III”, em 2010, este forçado pela crise de 2008 e 2009, no qual estabeleceu novas regras (MENDES, 2013).

Os principais objetivos da Basileia III são: melhorar a capacidade dos bancos de absorver choques decorrentes de estresse financeiro e econômico, qualquer que seja a fonte causadora; aprimorar as práticas de gestão e governança de riscos e fortalecer a transparência e as práticas de *disclosure*, “divulgação”, conforme descreve a Empresa de Auditoria *Price Water House Coopers* (PWC, 2013).

Devido as recorrentes crises que atingiu o mercado mundial financeiro e obviamente o Brasil, reforçado pelos acordãos da Basileia, principalmente após a Basileia III, houve forte interesse das instituições financeiras em estudos relacionados com o sistema e pontuação de crédito, com abordagem nas áreas de matemática e estatística e mais recentemente a computacional.

Esses sistemas trouxeram contribuição para o desenvolvimento de modelos de *credit scoring*, tornando sua utilização mais prática e fácil, sendo comum encontrar técnicas que atribuem pontuação ou escala de classificação quanto ao risco oferecido pelo cliente visando maximizar o lucro esperado pelas instituições financeiras (ABDOU e POINTON, 2011).

Como já exposto, a alteração do mercado financeiro e sua tentativa de prever o “*default*” através de técnicas de *credit scoring*, tem utilizado diferentes meios, o que torna relevante uma revisão das técnicas utilizadas. Neste sentido, Abdou e Pointon (2011), relatam que há poucos livros sobre o assunto, por ser relativamente recente, porém, descrevem o crescimento de estudos na área, com divulgação por meio de publicações em periódicos internacionais.

García *et al.* (2014), apresentaram relevante trabalho de revisão, publicado no *Journal of Intelligent Information Systems*, no qual foi efetuado uma pesquisa nos documentos publicados entre 2000 a 2013, com volume superior a 140 artigos. O objetivo da pesquisa foi “estudar como os experimentos foram projetados e os resultados validados no campo do risco de crédito e da previsão de falência corporativa”.

Os autores destacam a importância de ao menos ter 4 (quatro) fases de forma bem definida para tirar conclusões com boa fundamentação dos resultados, sendo: dados experimentais, as técnicas de divisão de dados, a avaliação de desempenho e os testes estatísticos de significância.

Quanto aos dados experimentais, os autores identificam algumas deficiências: Primeiro a limitação da base pública de dados, segundo o número reduzido de dados da amostra, o que pode aumentar a variação dos resultados, terceiro a utilização de um banco de dados exclusivo e único, aumentando a cautela quando das conclusões. Referente a divisão dos dados, identificou-se a falta de especificação da técnica empregada.

No que se refere avaliação de desempenho, a vasta maioria de artigos utilizou a precisão ou a taxa de erro, mesmo com dados desequilibrados de classe e diferentes custos de classificação errada. Os autores, descrevem ainda, que a preocupação do comportamento tendencioso da precisão pode induzir os pesquisadores a conclusões enganosas do modelo de predição.

Por fim, relativo aos testes estatísticos de significância, os autores descrevem que os mesmos não são muito frequentes. Assim no desenvolvimento do trabalho verificaram que alguns estudos apresentaram apenas o desvio-padrão e ausência do teste de hipóteses, sendo que outros aplicaram apenas um teste paramétrico sem verificar a normalidade dos dados.

Em outro trabalho relevante, Louzada *et al.* (2016), intitulado “*Classification methods applied to credit scoring: Systematic review and overall comparison*”, publicado no periódico *Surveys in Operations Research and Management Science*, realizaram uma revisão histórica de mais de 20 anos de pesquisa, incluindo 187 artigos entre os anos de 1992 a 2015.

Esse artigo teve como objetivo apresentar uma revisão sistemática da literatura relativa à teoria e aplicação de técnicas de classificação binária para análise financeira de pontuação de crédito.

Quanto a revisão, os autores observaram que o mais comum quanto aos objetivos foi de propor um novo método de classificação de *credit scoring*, principalmente com técnicas híbridas buscando desempenhos elevados.

Nos artigos, estudos de Louzada *et al.* (2016), as técnicas mais evidenciadas foram: Redes Neurais Artificiais (RNA), Máquina de Suporte Vetor, “*Support Vector Machines*” (SVM), sistemas híbridos e técnicas combinadas. As técnicas de Regressão Logística (RL),

Árvores de Decisão, “*Decision Tree*” (DT) e também RNA são bastante utilizadas em comparações com outras, sendo técnicas que devem ser superadas.

Como destaque, SVM apresentou-se como uma técnica de alto percentual preditivo e baixo esforço computacional, quando comparado a outras técnicas. Esses autores, destacam ainda, a preocupação dos cuidados com a interpretação dos resultados versus a validação dos dados e os tipos de erros que estão sujeitos.

O estudo de Prado *et al.* (2016), sob o título “*Multivariate analysis of credit risk and bankruptcy research data: a bibliometric study involving different knowledge fields (1968–2014)*”, objetivou identificar e descrever o uso de técnicas multivariadas de análise de dados em pesquisas sobre risco de crédito e falência.

A pesquisa foi realizada por meio de estudo bibliométrico, com dados coletados em publicações indexadas à base de dados da *Web of Science da Thomson Reuters* entre 1968 e 2014. Os resultados, desde o início da década de 1990, apontaram um crescimento das técnicas computacionais como RNA, transformando-se em técnica relevante, superar as técnicas estatísticas.

Ainda Prado *et al.* (2016), descrevem que as técnicas estatísticas de AD e RL, ainda são frequentemente usadas em pesquisas, havendo elevado número de citações envolvendo as 3 (três) técnicas: RNA, AD e RL, sendo destacado o crescimento de modelos híbridos.

A análise de crédito é um campo multidisciplinar, envolvendo diversas áreas do conhecimento como economia, negócios, gestão, engenharia e estatística, entre outras. Complementam, os autores, referindo-se a relevância do número crescente de publicações sobre esse assunto após crise de 2008, tendo praticamente dobrado o número de publicações em 2009.

Como questão de pesquisa, busca-se responder: Quais as principais técnicas utilizadas para construção de modelos de previsão de risco, de crédito para as pequenas e médias empresas (PME)?

No intuito de satisfazer a questão de pesquisa, estabeleceu-se como objetivo geral selecionar um portfólio de artigos relacionados com a questão de pesquisa. Especificamente, sobre a ótica do Knowledge Development Process – Constructivist (ProKnow-C), foram definidos os seguintes objetivos: (A) Selecionar o portfólio bibliográfico voltado a questão da pesquisa; (B) Realizar análise bibliométrica para identificar as técnicas de previsão de

risco, acurácia das mesmas, autores e periódicos que são destaques no período abrangido por este estudo.

Esta pesquisa tem característica exploratória, definida por Beuren (2014), como aquela que busca conhecer o assunto com maior profundidade. Complementa Gil, (2008), dizendo que a pesquisa exploratória pode ser classificada como bibliográfica.

Este trabalho está dividido em 4 (quatro) partes: Metodologia para seleção do portfólio, Resultados, Considerações Finais e Referências Bibliográficas, distribuídas nas seções 2, 3, 4 e 5 respectivamente.

2 METODOLOGIA PARA SELEÇÃO DO PORTFÓLIO

A seleção do material para realização da pesquisa, a fim de fornecer o embasamento teórico necessário e profundidade que este trabalho se propôs, utilizou-se o método Knowledge Development Process – Constructivist (Proknow-C). Entendem Afonso et al. (2011), que o método ProKnow-C se destaca na construção do conhecimento por proporcionar procedimento rigoroso, com redução da aleatoriedade e subjetividade do processo de revisão bibliográfica. Ainda Ensslin (2013), diz que o método já está consolidado, caracterizado por apresentar um processo para mapeamento do conhecimento, conforme percepção do pesquisador.

2.1 PERÍODO DE ESTUDO

O período estudado foi delimitado entre 2008 e 2016, por entender que dentro deste espaço- tempo estão registrados os principais impactos creditícios com repercussão nacional e internacional, dando uma visão geral da importância do crédito e de sua análise em termos globais. São eles: na esfera internacional, a explosão da Bolha Americana (subprime), com seu cume em 2008; e na esfera nacional, a forte recessão que atingiu o Brasil em 2013.

2.2 FONTES DA PESQUISA

Os sítios utilizados para busca de trabalhos foram: Web of Science e Scopus. Estes sítios correspondem a uma base de 1277 trabalhos de publicação nacional e internacional, o que motivou pela utilização dos mesmos na pesquisa.

2.3 CRITÉRIOS DE PESQUISA

Para retorno dos artigos foram definidos 2 eixos de pesquisa e 3 palavras-chave para cada eixo: Eixo (E) 1: Credit Analysis com as palavras-chave (PC) “Credit Scoring”, “Credit Risk” e “Techniques”. Eixo 2: Enterprice, com as palavras-chave (PC) “Prediction Models”, “Small and Medium-sized Enterprises” e “Variables”, relacionados com o tema da pesquisa, conforme descrito na Figura 1.

Na oportunidade da pesquisa, utilizou-se (“ ”) no início e final das palavras-chave, visando dar maior objetividade para composição do portfólio.

Com o cruzamento dos eixos de pesquisa com as palavras-chave, obteve-se 9 combinações. A Figura 1, apresenta a sua montagem, os conectores e o status do filtro de pesquisa, utilizados nos sites de pesquisa.

| Eixos | <i>Credit Analysis</i> | <i>Enterprices</i> | Conector | Filtro |
|--------------|-------------------------------|---|-----------------|-------------------|
| PC 1 | “ <i>Credit Scoring</i> ” | “ <i>Prediction Models</i> ” | | <i>All fields</i> |
| PC 2 | “ <i>Credit Risk</i> ” | “ <i>Small and Medium-sized Enterprises</i> ” | <i>And</i> | <i>All fields</i> |
| PC 3 | “ <i>Techniques</i> ” | “ <i>Variables</i> ” | <i>And</i> | <i>All fields</i> |

Fonte: Elaborado pelos autores
Figura 1 - Eixos *versus* Palavras-Chave

2.4 FILTRO DO PORTFÓLIO BIBLIOGRÁFICO

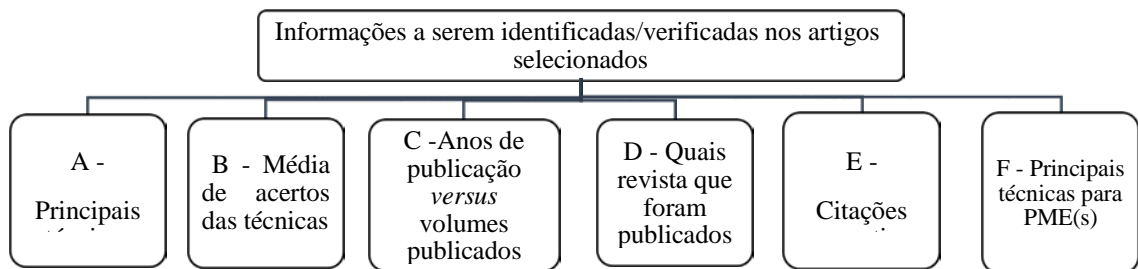
Para seleção dos trabalhos, utilizou-se de 4 fases distintas, sendo a 1ª fase voltada a observar quais dos trabalhos tem relação com o tema e área da pesquisa, sendo utilizado como base, o título do trabalho. A segunda fase, leitura dos resumos “abstracts”, visa selecionar artigos que tivessem relação com as técnicas de predição.

Posteriormente, na fase 3 em planilha do Excel e nos filtros das próprias páginas de pesquisa, foram ordenados de forma decrescente pelos números de citações, no qual foi selecionado apenas artigos que apresentaram número igual ou superior de 4 citações.

Para desenvolvimento das três fases descritas anteriormente, foram utilizados os Softwares: Mendeley, EndNote e Excel. Por fim, na fase 4, ocorreu a leitura completa dos trabalhos, com objetivo de analisar o conteúdo e classificá-los conforme aderência ao tema de pesquisa.

2.5 ANÁLISE BIBLIOMÉTRICA

De posse do portfólio selecionado, por meio do método Proknow-C, passou-se a efetuar uma revisão bibliométrica direcionada, com o objetivo de quantificar e identificar as características do portfólio. As questões utilizadas para o levantamento e quantificação dos dados foram: A) Principais técnicas, B) Média de acertos das técnicas, C) Anos de publicação versus volumes publicados, D) Quais revistas que foram publicados, E) Citações por artigo e F) Principais técnicas para PME(S), apresentadas na Figura 2.



Fonte: Elaborado pelos autores
 Figura 2 - Questões pesquisadas nos artigos do portfólio

3 RESULTADOS

3.1 SELEÇÃO DO PORTFÓLIO BIBLIOGRÁFICO VOLTADO A QUESTÃO DA PESQUISA

Com a aplicação das 9 combinações, utilizados nos sites de pesquisa, *Web of Science e Scopus*, dentro do período de 2008 a 2016, com utilização do *Proknow-C*, obteve-se o retorno de 1277 artigos.

A Figura 3 apresenta o volume de artigos encontrados em cada fase da seleção e sua forma de filtragem.

| Fases e número de artigos selecionados | | Ordem |
|--|------|---|
| Base inicial | 1277 | <i>Web of Science e Scopus</i> |
| Eliminação dos artigos duplicados | 369 | Aplicativo <i>Me ndelev</i> |
| Fase 1 | 908 | Títulos |
| Fase 2 | 372 | <i>Abstract</i> |
| Fase 3 | 69 | Número de citações igual ou maior que 4 |
| Fase 4 | 20 | Seleção para leitura completa |

Fonte: Elaborado pelos autores

Figura 3 – Apresentação das fases, números e ordem dos artigos pesquisados.

Conforme ilustra a Figura 3, os artigos selecionados, pós leitura completa, que compuseram o portfólio final, totalizaram 20, correspondendo a 2,20% do total de artigos selecionados na fase

1. A figura 4 apresenta a relação dos artigos utilizados neste trabalho, diferenciadas pelo tipo de empresa (Emp.): Geral (todos os tamanhos) e PME para pequenas e médias empresas.

| Nº | Artigo | Tipo | Ano |
|----|--|-------|------|
| 1 | Predicting China's SME Credit Risk in Supply Chain Financing by Logistic Regression, Artificial Neural Network and Hybrid Models | PME | 2016 |
| 2 | A genetic algorithm approach for {SMEs} bankruptcy prediction: Empirical evidence from Italy | PME | 2014 |
| 3 | Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises | PME | 2013 |
| 4 | Credit scoring models for the microfinance industry using neural networks: Evidence from Peru | PME | 2013 |
| 5 | Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples - Evidence from the Chinese hotel industry | Geral | 2012 |
| 6 | A Bayesian latent variable model with classification and regression tree approach for behavior and credit scoring | Geral | 2012 |
| 7 | Exploring the behaviour of base classifiers in credit scoring ensembles | Geral | 2012 |
| 8 | Decision tree-based technology credit scoring for start-up firms: Korean case | PME | 2012 |
| 9 | Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time | Geral | 2012 |
| 10 | Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis | Geral | 2011 |
| 11 | A support vector machine-based model for detecting top management fraud | Geral | 2011 |
| 12 | Could traditional financial indicators predict the default of small and medium-sized enterprises? | PME | 2011 |
| 13 | Credit rating prediction using Ant Colony Optimization | PME | 2010 |
| 14 | Predicting Korean lodging firm failures: An artificial neural network model along with a logistic regression model | Geral | 2010 |
| 15 | Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods | Geral | 2010 |
| 16 | Support vector machines for default prediction of SMEs based on technology credit | PME | 2010 |
| 17 | Predicting default of Russian SMEs on the basis of financial and non-financial variables | PME | 2009 |
| 18 | Using Economic-Financial Ratios for Small Enterprise Default Prediction Modeling: an Empirical Analysis | PME | 2008 |
| 19 | Fraudulent financial reporting detection and business failure prediction models: A comparison | PME | 2008 |
| 20 | Feature selection to diagnose a business crisis by using a real GA-based support vector machine: An empirical study | Geral | 2008 |

Fonte: Elaborado pelos autores
Figura 4 - Relação dos 20 artigos selecionados

Para confirmar a aderência das palavras-chave, conforme *Proknow-C*, foram selecionados 2 artigos de forma aleatória que são: “A *support vector machine-based model*

for detecting top management fraud e Support vector machines for default prediction of SMEs based on technology credit”, os quais mostraram boa aderência das palavras-chave.

3.2 APRESENTAÇÃO DOS RESULTADOS DA ANÁLISE BIBLIOMÉTRICA

Inicialmente foram codificadas as técnicas de predição de *credit scoring*, conforme apresentado na Figura 5. Alguns artigos apresentam técnicas combinadas ou adaptadas a outros métodos, objetivando melhorar a acurácia do modelo.

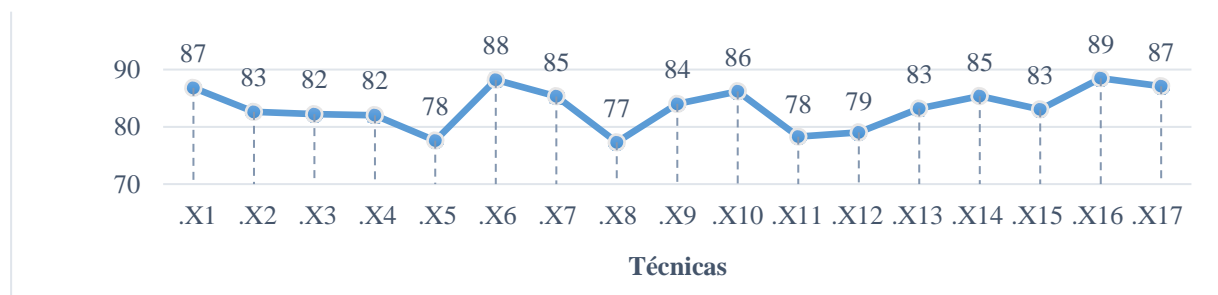
Para uma melhor leitura da técnica principal, as composições ou adaptações ocorridas, foram agrupadas em um único código (Cód), objetivando facilitar a compreensão e análise das mesmas. As técnicas agrupadas, correspondem aos Cód: X1, X4, X5, X6 e X 14.

| | | | |
|-----|--|-----|---|
| X1 | SVM support vector machines e GA-SVM genetic algorithm e support vector machines | X10 | Principles from ant colony optimization (AntMiner+) |
| X2 | RNA artificial neural network | X11 | MARS multivariate adaptive regression splines |
| X3 | AD discriminant analysis | X12 | BNP bayesian |
| X4 | RL logistic regression e model 2 | X13 | Majority vote |
| X5 | RBF kernel e RBF radial basis function | X14 | Decision tree: BFP; C4.5 |
| X6 | (k-NN) e (1-NN) nearest neighbors | X15 | MLP multilayer perceptron |
| X7 | (CART) Classification and regression tree | X16 | Hybrid two-stage (LR e RNA) model III. |
| X8 | Kohonen map and trajectory design | X17 | Genetic algorithms GA(s) |
| X09 | Cox | | |

Fonte: Elaborado pelos autores

Figura 5 - Descrição das técnicas e agrupamento da semelhantes

Para identificar as técnicas de previsão de risco, utilizadas na formação de *credit scoring*, respondendo as questões A e B, foram tabulados os dados de predição encontrados nos 20 artigos selecionados. Para apresentação da acurácia, foi calculado a média simples das técnicas, conforme descrito na fase 4, as quais são apresentadas na Figura 6.

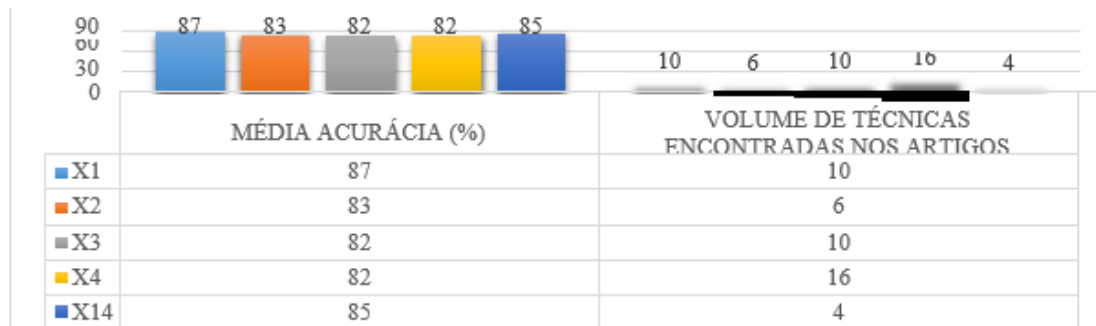


Fonte: Elaborado pelos autores

Figura 6 - Média de predição das técnicas

Importante salientar, que são utilizadas mais de uma técnica por artigo, justificando o volume de técnicas apresentadas. Conforme destacado na Figura 6, as técnicas X6 e X16, aparecem com acurácia próxima a 90%. Além disso, cabe ressaltar que dentro dos 20 artigos selecionados as técnicas respectivas de (k-NN) e (1-NN) *nearest neighbors* e *Hybrid two-stage* (LR e RNA) *model III* foram aplicadas uma única vez, o que dificulta a análise de sua performance em diferentes cenários.

Devido a pesquisa, na seleção dos 20 artigos, não ter encontrado repetições de aplicação das técnicas em destaque na Figura 6, foi elaborado um gráfico, apresentado na Figura 7, onde destaca as técnicas que figuraram 4 ou mais vezes, no total dos artigos que compuseram este estudo. A figura 7 apresenta ainda o volume (quantidade) de vezes que foram utilizadas na predição de risco de crédito e falência.



Fonte: Elaborado pelos autores

Figura 7 - Predição e volume de técnicas encontradas nos modelos nos 20 artigos selecionados

Em relação as técnicas mais utilizadas, observa-se que X4 “Regressão Logística” é destacadamente a mais utilizada. Isso ocorre, porque essa técnica está consolidada na teoria e apresentar-se como uma boa ferramenta de predição. A Regressão Logística é uma das técnicas estatísticas mais utilizadas para elaboração de modelos de *credit scoring*, nas quais as variáveis dependentes são binárias (LUO *et al.*, 2016).

Observou-se, que esta técnica é bastante utilizada como comparativo de eficiência, “parâmetro”, para outras técnicas. Posteriormente, com 10 utilizações seguem X1 e X3, e com 6 e 4 utilizações respectivamente, aparecem as técnicas descritas em X2 e X14. Todas alcançando percentuais de acurácia acima de 80%

Na sequência das questões a responder, conforme informações levantadas no portfólio, em resposta a questão C: anos de publicação *versus* volume de artigos publicados, observa-se que no ano de 2012 atingiu volume máximo, conforme Figura 8.

Nos anos anteriores de 2010 e 2011, mesmo para um portfólio reduzido de 20 artigos, apresenta uma variação entre 3 e 4 artigos publicados, superior aos anos de 2008 e 2009, período que coincide com a expansão da crise internacional, denominada de *subprime*, conforme descrito por (PRATES e FARHI, 2009).

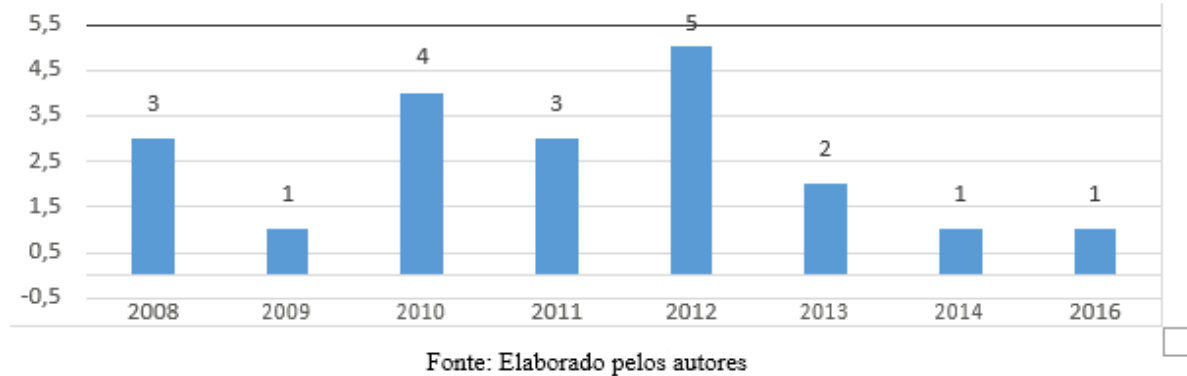


Figura 8 - Anos das publicações *versus* volume de documentos publicados

Em resposta ao questionamento D, a Figura 8 apresenta as revistas que os artigos foram publicados. Interessante observar que se destaca a revista *Expert Systems With Applications* com 6 trabalhos, seguida da *Knowledge Based Systems e European journal of Operational Research*, as quais apresentaram 2 artigos publicados, em cada revista, dentro do portfólio, sendo apresentado na Figura 9.

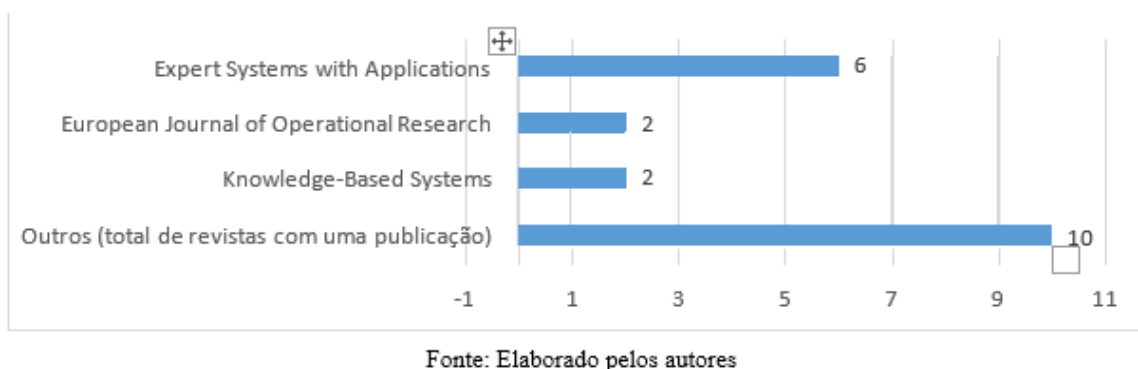
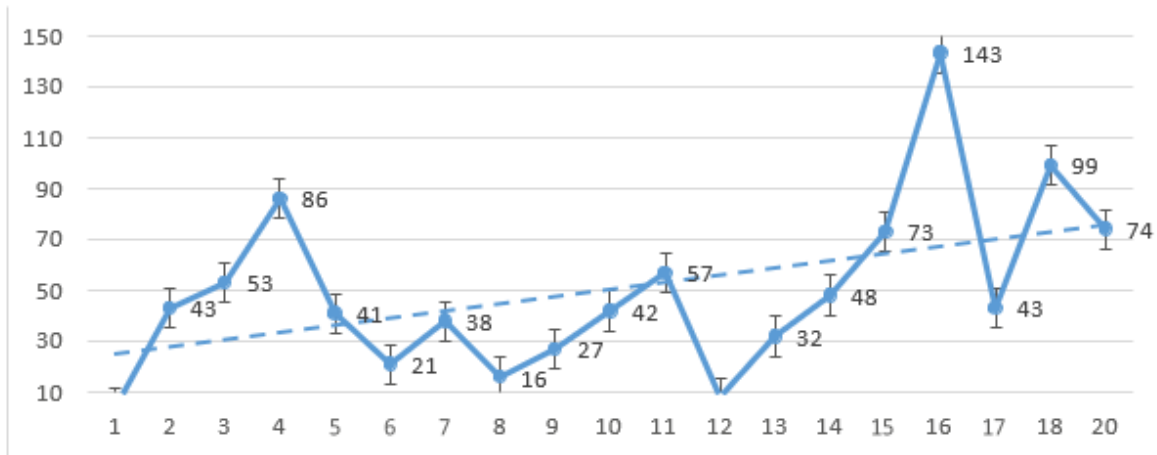


Figura 9 – Número de publicações por revista que compuseram o portfólio selecionado

Na sequência, em resposta a questão E, citações autorais por artigo, fez-se uma busca no site do Google Acadêmico para verificar o volume de citações por artigo.

Conforme ilustra a Figura 10, dentre os 20 artigos, observa-se um maior volume de citações para o artigo intitulado *Support vector machines for default prediction of SMEs based on technology credit*, apresentando 143 citações, publicado em 2010 na revista *European Journal of Operational Research*.

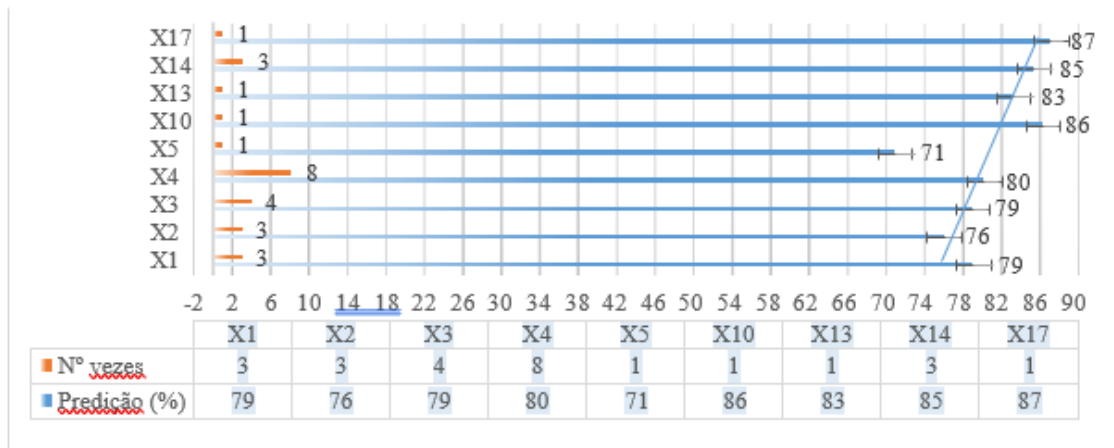


Fonte: Elaborado pelos autores

Figura 10 - Apresenta o número de citações de cada artigo que compôs o portfólio

Finalmente, para responder à questão F: principais técnicas e quantidade de vezes que foram utilizadas, dentro do portfólio, exclusivamente para as pequenas e médias empresas, “*small and medium-sized enterprises*”, elaborou-se o gráfico representado na Figura 11.

Este gráfico teve a preocupação de ilustrar as técnicas que foram utilizadas nos artigos, que direcionaram esforços para predição ou falências, extraindo-se 11 artigos do total do portfólio e 9 técnicas.



Fonte: elaborado pelos autores

Figura 11 - Apresenta a acurácia das técnicas e o número de vezes que foram utilizadas dentro do portfólio dos 11 artigos direcionados para as PME(s).

Observa-se que a técnica que mais figurou, no volume dos 11 artigos direcionados as PME(s), foi X4, utilizada em 8 artigos e com predição média de 80%, sendo que a técnica que teve melhor predição foi X17, porém, figurou apenas em 1 artigo.

4 CONSIDERAÇÕES FINAIS

Este artigo, preocupou-se em apresentar uma revisão sistemática da literatura sobre as técnicas em evidência, recomendada pela literatura para modelagem de *credit scoring*. Este trabalho não busca fazer um apanhado geral e definitivo sobre o tema, mas observar, em determinado momento, de forma selecionada, os estudos que estão acontecendo sobre o tema. As pesquisas sobre, técnicas de predição, tiveram maior volume em decorrência de, ao menos, 2 fatores: as crises internacionais e suas intensidades sobre o setor econômico financeiro e também as influências relacionadas a emissão de normas da Basileia, principalmente a II e III, as quais incorporaram preocupações com o equilíbrio financeiro e as medidas para melhor controle da inadimplência (PRATES e FARHI, 2009; LOUZADA *et al.*, 2016).

Com esta revisão, foi possível verificar a utilização das diversas técnicas, bem como suas variações, alterações e aprimoramentos, principalmente AD, RL, RNA e SVM. Dentro do portfólio, foram as técnicas mais utilizadas para predição de risco e *credit scoring*, para empresas de pequeno e médio porte.

Quando comparado a Figura 11 com a Figura 06, observou-se que a média da acurácia, principalmente das técnicas X1, X2, X3 e X4, são menores para as PME(s),

permanecendo a técnica X14 com a mesma predição. As causas dessa diferença, apresentada nas técnicas acima descritas, podem estar na definição das variáveis, uma vez que as PME(s), tem dificuldades de geração de indicadores contábeis dirigidos para gestão, por terem menos obrigações legais em matéria de divulgação (BERTI, 2012; CIAMPI e GORDINI, 2013). Porém, essa variação não foi abordada especificamente por este artigo.

Vale ressaltar que a literatura não apresenta uma fórmula única ou técnica consensual ou superioridade de uma técnica sobre outra, mesmo porque as bases de dados e as situações alteram-se de instituição para instituição na busca de predição entre bons e maus pagadores. Outro fator importante na escolha da técnica é a facilidade de alteração do modelo em decorrência de mudanças de cenário ou faixas de atuação.

Verificou-se evidências de estabilidade nas 4 técnicas apresentadas AD, RL, RNA e SVM e suas variações ou complementações. As técnicas de AD, RL e RNA apareceram fortemente dentro dos trabalhos servindo como base para avaliação de outras técnicas e suas eficiências comparativamente a estas. Estas técnicas, entre outras, também foram destacadas no trabalho de (PRADO *et al.*, 2016; LOUZADA, 2016).

A escolha da técnica está relacionada às ferramentas que o analista/especialista possui, no entanto, o banco de dados pode contribuir nesta decisão, uma vez que ele é a matéria-prima da técnica, portanto, sua forma de tratamento, quando da manipulação, torna-se vital para a acurácia e utilização adequada da mesma.

Quanto a geração da informação, obtida da compilação dos dados levantados, verifica-se a busca por técnicas que possam retornar um resultado cada vez mais acurado, buscando a objetividade na decisão, na busca da reduzindo a subjetividade e as intempéries do humor ou grau de conhecimento do especialista.

De acordo com o objetivo deste trabalho a revisão alcançou sua finalidade na observação de um conjunto de artigos selecionados, identificando as técnicas que foram utilizadas nos mesmos, no período de 2008 a 2016, para as empresas de pequeno e médio porte.

Deixa-se, como sugestão de estudos posteriores, a aplicação de técnicas para predição de risco de crédito para pequenas e médias empresas utilizando dados auditáveis e gerenciais.

REFERÊNCIAS

- ABDOU, H. A.; POINTON, J.** *Credit Scoring, Statistical Techniques And Evaluation Criteria : A Review Of The Literature.* Journal of Small Business Management. Vol. 88, n. 6, p. 59–88, 2011.
- AFONSO, M. H. F.; SOUZA, J. V.; ENSSLIN, S. R.; ENSSLIN, L.** *Como construir conhecimento sobre o tema de pesquisa? Aplicação do processo Proknow-C na busca de literatura sobre avaliação do desenvolvimento sustentável.* Revista de Gestão Social e Ambiental. Vol. 5, n. 2, p. 47-62, 2011.
- BERTI, A.** *Consultoria e Diagnóstico Empresarial.* 2. ed. Curitiba: Juruá, 2012.
- BEUREN, I. M.** *Como elaborar trabalhos monográficos em contabilidade: teoria e prática.* - 3. ed. 9. Reimpr. São Paulo: Atlas, 2014.
- BLANCO, A.; MEJÍAS, P.; LARA, J.; RAYO, S.** *Credit scoring models for the microfinance industry using neural networks: Evidence from Peru.* Expert Systems With Applications. Vol. 40, n. 1, p. 356–364, 2013.
- CHEN, L. H.; HSIAO, H. D.** *Feature selection to diagnose a business crisis by using a real GA-based support vector machine: An empirical study.* Expert Systems with Applications. Vol. 35, n. 3, p. 1145–1155, 2008.
- CIAMPI, F.; GORDINI, N.** *Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises.* Journal Of Small Business Management. Vol.. 51, n. 1, p. 23–45, 2013.
- CIAMPI, F.; GORDINI, N.** *Using Economic-Financial Ratios for Small Enterprise Default Prediction Modeling: An Empirical Analysis.* Oxford Business & Economics Conference Program. n. January 2008, 2008.
- DIEESE.** *Departamento Intersindical de Estatística e Estudos Socioeconômicos.* Evolução do crédito na economia brasileira, 2008-2013. n. 135, 2014.

DURAND, D. *Risk elements in consumer instalment financing, in: National Bureau of Economics, New York. 1941.*

DU JARDIN, P.; SÉVERIN, E. *Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time.* European Journal of Operational Research, Vol. 221, n. 2, p. 378–396, 2012.

ENSSLIN, L. *Processo de Investigação e Análise Bibliométrica : Avaliação da Qualidade dos Serviços Bancários Research Process and Bibliometric Analysis: Bank Service Quality Assessment.* p. 325–349, 2013.

GARCÍA, V.; MARQUÉS, A. I.; SÁNCHEZ, J. S. *An insight into the experimental design for credit risk and corporate bankruptcy prediction systems.* Journal of Intelligent Information Systems. 2014.

GIL, A. C. *Como elaborar projetos de pesquisa.* 4. ed. São Paulo: Atlas. 2008.