



DESEMPENHO FINANCEIRO COMO CRITÉRIO DE DECISÃO PARA SELEÇÃO MODELOS DE ANÁLISE E CONCESSÃO DE CRÉDITO

1- Rodrigo Alves Silva*

Programa de Pós-doutorado em Economia Aplicada - Departamento de Economia, Administração e Sociologia.
Universidade de São Paulo, Brasil.
ralves08@usp.br
<http://lattes.cnpq.br/5156569312693322>

2- Anderson Ara

Departamento de Estatística, Universidade Federal da Bahia (UFBA), Brasil.
anderson.ara@ufba.br
<http://lattes.cnpq.br/8916772290938469>

3- Evandro Marcos Saidel Ribeiro

Programa de Pós-Graduação em Administração de Organizações – PPGAO, FEA-RP/USP – Universidade de São Paulo, Brasil.
esaidel@usp.br
<http://lattes.cnpq.br/1513305763913243>

Diego Maganhotto Coraiola – Editor Geral

Editor responsável pela submissão:

Diego Maganhotto Coraiola.

Artigo analisado via processo de revisão duplo cego (*Double-blind*).

Recebido em: 24/01/2017

Aprovado em: 22/05/2017

Última Alteração: 09/08/2017

* Contato Principal: DAGEE – Departamento Acadêmico de Economia e Gestão. UTFPR – Universidade Tecnológica Federal do Paraná. Avenida Sete de Setembro, 3165 – Rebouças – Curitiba – PR. CEP: 80230-901.

DESEMPENHO FINANCEIRO COMO CRITÉRIO DE DECISÃO PARA SELEÇÃO MODELOS DE ANÁLISE E CONCESSÃO DE CRÉDITO

RESUMO

O presente artigo tem por objetivo apresentar a importância do emprego de métricas financeiras na tomada de decisão de seleção de modelos de *credit scoring*. Para tanto, considerando um sistema de aprovação automática, este artigo empregou métricas de desempenho financeiro sobre as carteiras teóricas geradas por modelos de *credit scoring* baseados em sete das principais técnicas de aprendizagem estatística utilizadas para tratamento dessa problemática. Os modelos foram ajustados aos dados da base *German Credit Data Set* e os resultados analisados com base em quatro métricas: acurácia total, custo por erro, retorno ajustado ao risco de capital e índice de Sharpe. Os resultados apontam que a acurácia total, amplamente empregada como critério de seleção de modelos de *credit scoring*, é incapaz de selecionar o modelo mais rentável para a firma, indicando a necessidade de incorporar métricas financeiras ao processo de seleção de modelos de *credit scoring*.

Palavras-chave

Risco de crédito; Seleção de modelos; Aprendizagem estatística.

FINANCIAL PERFORMANCE AS A DECISION CRITERION OF CREDIT SCORING MODELS SELECTION

ABSTRACT

This paper aims to show the importance of the use of financial metrics in decision-making of credit scoring models selection. In order to achieve such, we considered an automatic approval system approach and we carried out a performance analysis of the financial metrics on the theoretical portfolios generated by seven credit scoring models based on main statistical learning techniques. The models were estimated on German Credit dataset and the results were analyzed based on four metrics: total accuracy, error cost, risk adjusted return on capital and Sharpe index. The results show that total accuracy, widely used as a criterion for selecting credit scoring models, is unable to select the most profitable model for the company, indicating the need to incorporate financial metrics into the credit scoring model selection process.

Keywords

Credit risk; Model's selection; Statistical learning.

1 Introdução

Operações de crédito são caracterizadas pela troca de bens, serviços ou recursos no instante atual por uma promessa de pagamento futura. A probabilidade de não ocorrência desse pagamento no futuro e as consequentes perdas do emprestador são os fatores que geram a necessidade de desenvolvimento de modelos para análise e concessão de crédito.

Dada a quantidade de operações e o volume envolvido, a gestão de carteiras de crédito é um tema recorrente em finanças corporativas e bancárias, sendo, tais carteiras, foco de constantes esforços de otimização (Lu, Huang, Ching & Siu, 2013). Uma das principais atividades da gestão da carteira de crédito é a previsão de insolvência, também conhecida como previsão de default de crédito. Entende-se como previsão de *default* o esforço de avaliação pré-concessão de crédito realizado pelo emprestador com o objetivo de antever solicitantes que não conseguirão honrar com os compromissos contratados em caso de concessão, ficando em default.

Para Duan e Shrestha (2011) dadas as consequências trazidas pelo *default* às firmas, seus fornecedores, clientes, funcionários, dentre outros *stakeholders*, um dos principais componentes da gestão de carteiras de crédito é a previsão da insolvência ou *default*. Segundo os autores, ainda que a extensão dos danos à economia dependa do tamanho e das características sistemáticas do tomador em *default*, as perdas oriundas do *default* são potencialmente danosas tanto em pequenas empresas sem qualquer implicação sistemática quanto em grandes corporações.

Iscoe, Kreinin, Mausser e Romanko (2012) indicam que as decisões de concessão e a gestão de portfólio de crédito são baseadas em informações de modelos quantitativos de previsão e otimização. Segundo Lu, *et al.* (2013) e García, Giménez e Guijarro (2013) esses modelos são dedicados à minimização das perdas provenientes dos diversos riscos das operações de crédito, tais como: risco de mercado, risco de crédito, risco de liquidez, risco operacional e risco sistemático.

A probabilidade de ocorrência de mudanças na qualidade do tomador (risco de crédito) em uma operação isolada pode ser definida como a esperança de um tomador de crédito não honrar seus compromissos, gerando perdas à organização (García, Giménez & Guijarro, 2013). A estimativa dessa esperança servirá tanto para aprovar ou rejeitar solicitações de crédito, quanto para desenvolver estimativas do custo da operação da organização no mercado, do volume para coberturas (alocação de capital econômico) e dos prêmios pelo risco que serão embutidos nas taxas de retorno exigidas pelos fornecedores de crédito (Bluhm, Overbeck & Wagner, 2010).

Tomadores são organizados em classes (*ratings*) por semelhança quanto ao risco de *default*, sendo que o perfil da carteira será a soma ponderada do risco de cada *classe de rating* que compõe a carteira. Como parte do processo gerencial, os modelos de análise e concessão de crédito buscam atender aos anseios e políticas no que tange o perfil de carteira e de *exposure* decididos pelos fornecedores de crédito. Bluhm, Overbeck e Wagner (2010) apontam que políticas mais permissivas podem ter efeitos indesejáveis, pois podem aumentar as perdas oriundas do *default*. Em contrapartida, políticas restritivas de crédito também são potencialmente danosas no sentido de que podem provocar insatisfação em clientes e menor volume de vendas, reduzindo a escala e a lucratividade.

O desenvolvimento tecnológico, as características mercadológicas e a busca pelo aprimoramento da estrutura prudencial das instituições financeiras, com destaque para os acordos da Basileia I, II e III, têm incentivado o desenvolvimento de modelos internos para a mensuração, tratamento e gestão do risco de crédito (Bluhm, Overbeck & Wagner, 2010). Todavia, as métricas empregadas para a seleção de modelos geralmente consideram o nível de *acurácia* como medida comparativa de eficiência. Estudos como os de García, Giménez e Guijarro (2013) Xiao, Xiao e Wang (2016), Iscoe, *et al.* (2012), Pascual, Martínez e Alamillos (2014), Louzada, Ferreira-Silva e Diniz (2012) e Brown e Mues (2012) empregam a acurácia total como medida da eficiência do modelo.

Entretanto, é importante considerar que a otimização do crédito vem de um equilíbrio entre as perdas derivadas do *default* e das perdas derivadas de políticas de restrição de crédito. Como expressam Iscoe *et al.* (2012) os modelos devem oferecer proteção contra perdas sem, contudo, prejudicar o volume de negócios.

Partindo da premissa de que o desempenho da carteira não está associado apenas ao nível de acerto do modelo, mas também ao resultado financeiro oriundo da atividade de crédito, o presente estudo tem por objetivo testar o critério de decisão comumente empregado para seleção de modelos de *credit scoring*. Para tanto, a pesquisa compara o desempenho de modelos distintos através de métricas estatísticas e financeiras buscando verificar se modelos com maior acurácia total também apresentam resultados financeiros superiores.

Assim, a pesquisa visa contribuir para um novo paradigma de seleção de modelos de análise de risco de crédito útil a organizações financeiras e não financeiras que realizem operações de concessão de crédito, paradigma este voltado para a eficiência financeira dos modelos e não apenas para a eficiência estatística.

2 Classificadores em Análise e Concessão de Crédito

Como explicam Bluhm, Overbeck e Wagner (2010), o risco de crédito é mensurado como a esperança dos resultados das probabilidades dos possíveis estados da natureza associados à atividade. Tal mensuração pode ser realizada através da estimativa de três componentes: i) da esperança da probabilidade de insolvência para os clientes – DP (*Default Probability*); ii) da fração média da perda em caso de insolvência – LGD (*Loss given default*); e iii) da exposição total à insolvência do período – EAD (*Exposure at default*). Para calibrar a probabilidade de *default* os fornecedores de crédito utilizam modelos internos de *rating*, além de agências externas, tais como a Mood's Investors Services, a Standar & Poor's e a Fitch; ou ainda, valerem-se de dados de mercado, através de modelos como KMV¹, Credit Metrics, Credit risk+, dentre outros (Glantz & Kissell, 2014). Como observam Glantz e Kissell (2014), modelos de *credit scoring* podem ser empregados tanto para calibração das probabilidades de *default* das carteiras via modelos internos quanto para análise das solicitações, maximizando a importância dos mesmos.

A seleção das técnicas aplicadas ao problema de *credit scoring* é influenciada pelo perfil de pesquisa e de problema de decisão em análise. Como abordam Louzada, Ara e Fernandes (2016) as técnicas selecionadas podem buscar comparar técnicas tradicionais com novas técnicas de classificação, discussões conceituais, características de seleção, revisão de literatura, estudos de mensuração de performance, dentre outros objetivos.

Por meio de uma revisão sistemática, Louzada, Ara e Fernandes (2016) analisaram o uso de técnicas de *credit scoring* em diversos estudos publicados em quatro bases de dados que cobriram mais de 20 mil estudos disponíveis por meios eletrônicos. O estudo cobriu os anos de 1992 a 2015, sendo levantados 187 estudos especificamente voltados para técnicas de *credit scoring*. Em seus resultados, Louzada, Ara e Fernandes (2016) apontam um crescimento no número de pesquisas sobre o tema por ano e atribuem às novas técnicas e ao desenvolvimento tecnológico tal crescimento. Especificamente em estudos de comparação, como o proposto por este estudo, os autores indicam que as técnicas mais utilizadas são as técnicas de regressão logística, redes neurais, árvores e SVM – *Support Vector Machines*, como demonstra a Figura 1.

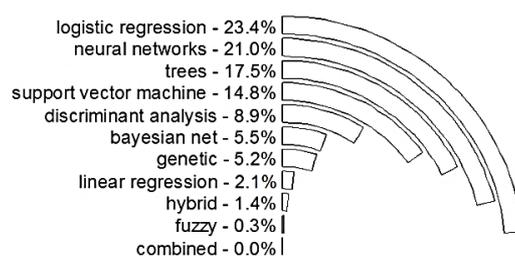


Figura 1. Uso de técnicas estatísticas em comparações de *credit scoring*
Fonte: Louzada, Ara e Fernandes (2016)

Como expressam Duan e Shrestha (2011), as técnicas mais utilizadas em modelos de *credit scoring* são a análise discriminante e a regressão logística. Em seu estudo comparativo os autores trabalham ainda outras três técnicas: redes neurais artificiais, SVM – *Support Vector Machine* e PIM – *Poisson Intensity Model*. Já Pascual, Martínéz e Alamillos (2013) sinalizam para o aumento do uso de redes bayesianas para o desenvolvimento de modelos de *credit scoring*, *credit ratings* dentre outras aplicações na área de crédito. Segundo os autores esta técnica tem se demonstrado poderosa para tratamento deste tipo de problemática.

Considerando tais indicativos, para o presente estudo foram comparados modelos baseados nas seguintes técnicas: análise discriminantes, regressão logísticas, redes bayesianas e SVM.

2.1 Métodos de Classificação Estudados

Nesta sessão serão apresentadas e discutidas as técnicas de classificação estudadas pela pesquisa. Maiores detalhamentos sobre as técnicas, bem como sobre o uso dos algoritmos para aplicação, podem ser observados em três principais bibliografias: Hair, Black, Babin e Anderson (2014), Nagarajan, Scutari e Lèbre (2013) e James, Witten, Hastie e Tibshirani, (2013).

Pode-se entender análise discriminante como uma técnica que tem por objetivo classificar uma entidade em um número finito de classes (grupos ou categorias) pré-estabelecidas com base em suas características (variáveis), assumindo que cada entidade pertence a apenas uma dessas classes (Hair *et al.*, 2014). A análise linear discriminante em sua forma funcional pode ser expressa como:

$$Z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n$$

Em que Z representa a variável predita, α é uma constante e $\beta_1 \dots \beta_n$, representam os coeficientes das variáveis e $x_1 \dots x_n$ são as variáveis discriminantes.

Um dos principais aspectos do emprego da análise discriminante é a sua semelhança aos modelos de regressão múltipla, o que facilita a interpretação e replicação, servindo também de base de comparação para análise de resultados de outras técnicas.

A regressão logística é uma técnica que considera um conjunto de variáveis explicativas $x = \{x_1, x_2, \dots, x_n\}$ e uma variável explicada binária $y = \{y_1, y_2\}$. A estimativa do modelo consiste em uma combinação linear de X e de uma transformação logit de Y. O modelo retorna em Y_a um valor dentro de um intervalo de 0 a 1 que pode ser interpretado como a probabilidade de Y_a pertencer ao grupo de interesse (Louzada, Ara, & Fernandes, 2016). Em sua forma funcional, a regressão logística pode ser expressa como:

$$Y_a = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

Em que $\beta_1 \dots \beta_n$, representam os coeficientes das variáveis e $x_1 \dots x_n$ são as variáveis independentes.

De acordo com Nagarajan, Scutari e Lèbre (2013) uma rede bayesiana é uma classe de modelos gráficos que permite construir uma representação de dependência probabilística entre um dado conjunto de variáveis aleatórias $X = \{X_1, X_2, \dots, X_p\}$ representada por um grafo direto e acíclico (DAG – *Direct acyclic graph*) $G = (V, A)$ no qual cada nó $V_i \in V$ corresponde a uma variável aleatória X_i . Os grafos apresentam a estrutura de dependências de uma tabela de probabilidade condicional que é obtida a partir de um processo de fatoração e que expressa a probabilidade de ocorrência de cada estado da natureza da variável de interesse, dada a ocorrência das variáveis predecessoras, como segue:

$$p(x_1, x_2, \dots, x_n) = \prod_{j=1}^n p[x_j | PAR_g(x_j)]$$

Em que PAR_g representa as variáveis *predecessoras* de x_i (também conhecidas como pais de x_i).

Como abordam Pascual, Martín e Alamillos (2013), há um grande número de tipos de redes bayesianas. Para o presente estudo serão abordadas as redes bayesianas Naive Bayes e KDB. Pascual, Martín e Alamillos (2013), ressaltam que a principal suposição do classificador Naive Bayes é a de que as variáveis (nós) são condicionalmente independentes dada a variável predita. Isto implica dizer que no Naive Bayes um nó filho não poderá ser pai de outro nó filho. Já na rede bayesiana kDB (*k-Dependence Bayesian Network*), ao contrário do Naive Bayes, existe uma relação de precedência na qual um filho é pai de outro filho.

Na figura 2.(a), não existe qualquer dependência (Naive Bayes ou kDB-0). Já na figura 2.(b) um (nó) filho é pai de, pelo menos, outro filho. Nesta estrutura o número 1 (kDB-1) significa que um filho que é pai de um outro filho. A estrutura kDB pode assumir então n relações de dependência, isto é, kDB-n.

Segundo Louzada e Ara (2012), para o caso estritamente discreto, a probabilidade à posteriori de um evento, dado o conjunto de variáveis, pode ser calculado pelo kDB a partir de:

$$P(Y = y_k | x_1, x_2, x_3, \dots, x_p) = \frac{P(Y = y_k) \prod_{i=1}^p P(x_i | \text{pais}(X_i), Y = y_k)}{\sum_j P(Y = y_j) \prod_{i=1}^p P(x_i | \text{pais}(X_i), Y = y_j)}$$

Outros classificadores que têm ganhado notoriedade na área de risco de crédito são os classificadores baseados no conceito de MMC – *Maximal Margin Classifier*. Os dois principais métodos derivados desta abordagem são o SVC (*Support Vector Classifier*) e o classificador de mesmo nome SVM (*Support Vector Machine*).

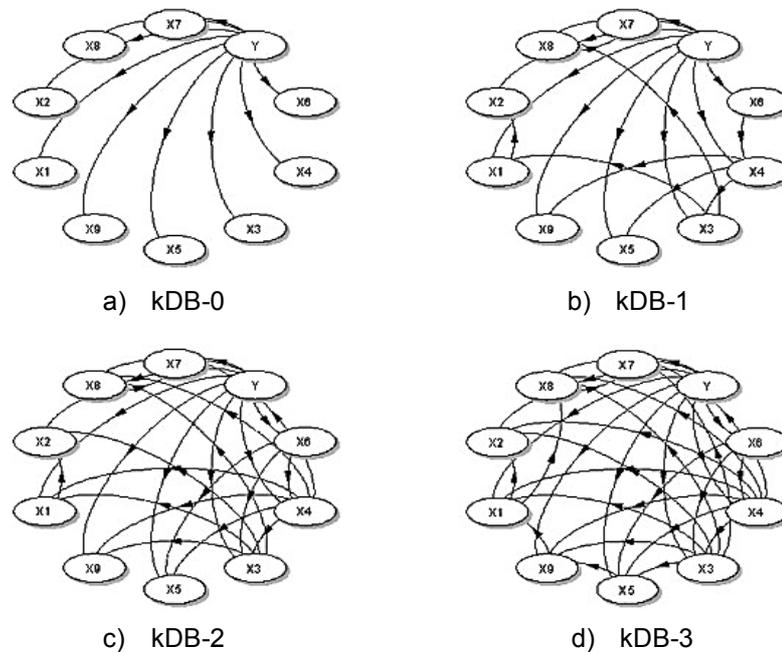


Figura 2. Estruturas de dependência em redes bayesianas
 Fonte: Louzada e Ara (2012, p.11585).

Em ambos os casos, a classificação está baseada no conceito de hiperplano². Considerando o conceito de hiperplano, na problemática de classificação – como é o caso dos problemas de previsão de *default* – existem muitos casos nos quais não é possível separar perfeitamente um conjunto de dados em duas (ou mais) classes pela observância de um máximo hiperplano linear. O SVC, também conhecido como SMC (*Soft Margin Classifier*), é um método que baseia a classificação no entendimento de que o máximo hiperplano não necessariamente é o melhor vetor de separação ou classificação das entidades existentes no espaço em questão (James *et al.*, 2013).

O método abre mão do máximo hiperplano e determina um hiperplano que seja tão menor quanto necessário para conseguir o máximo em termos de acerto nas classificações, o que é entendido como o ótimo hiperplano.

Assim como nos demais métodos de classificação baseados em um hiperplano, o SVC classifica a observação em função do lado do hiperplano ao qual ela se encontra, contudo, classificações erradas são inevitáveis. A classificação é obtida a partir de um conjunto de dados de treinamento no qual o algoritmo busca fornecer a melhor separação possível dentro do conjunto de dados. James *et al.* (2013), apontam que, para classificação em duas classes, busca-se a solução para o seguinte problema de otimização:

$$\text{Maximizar } M$$

$$\beta_0, \beta_1, \dots, \beta_p, \epsilon_0, \epsilon_1, \dots, \epsilon_n$$

Sujeito a:

$$\sum_{j=1}^p \beta_j^2 = 1;$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i),$$

$$\epsilon_i \geq 0, \sum_{i=1}^n \epsilon_i \leq C,$$

Em que M representa a margem, C é um parâmetro de ajuste não negativo, M é a largura da margem, devendo esta ser tão maior quanto possível; $\epsilon_1 \dots \epsilon_n$ são chamadas de variáveis de folga e têm a função de permitir que uma observação esteja no lado errado do hiperplano (ou da margem).

A classificação é obtida com base no sinal da função $f(x^*) = \beta_0 + \beta_1 x_1^* + \dots + \beta_p x_p^*$.

O *Support Vector Machine* – SVM – pode ser definido como uma extensão do SVC, sendo que sua principal diferença para o SVC é que o SVM permite a incorporação de estimativas não lineares em suas classificações (James *et al.*, 2013).

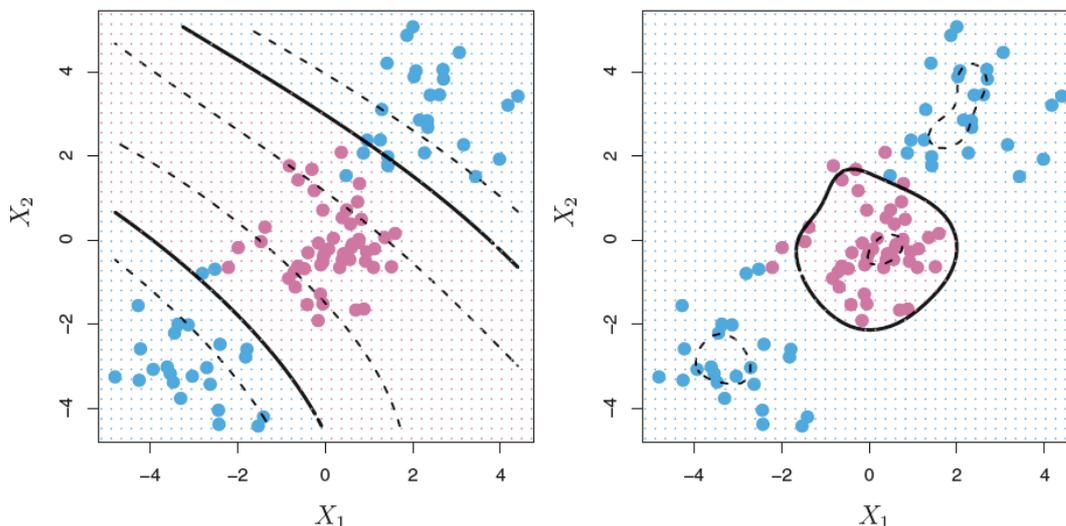


Figura 3. Comportamento não linear do SVM
Fonte: James *et al.* (2013, p. 350).

Considerando a Figura 3, observa-se este comportamento de forma bastante clara, tendo em vista que não há um hiperplano que separe linearmente as classes de uma forma eficiente. Como abordado por James *et al.* (2013), esse é um comportamento bastante comum nos dados reais.

O *Support Vector Machine* – SVM, incorpora ao SVC variáveis do tipo kernel (granulares), que são funções capazes de quantificar a similaridade entre duas observações. Considerando o SVC linear para duas observações (x_i, x_i') , como sendo dado por: $\langle x_i, x_i' \rangle = \sum_{j=1}^p x_{ij} x_{i'j}$. O SVC pode ser representado por:

$$f(x) = \beta_0 \sum_{i \in S} \alpha_i \langle x_{ij}, x_{i'j} \rangle$$

Em que S representa uma coleção de pontos de suporte.

Incorporando a variável kernel à função para duas variáveis (x_i, x_i') , esta poderia ser representada como:

$$K(x_i, x_i') = \left(1 + \sum_{j=1}^p x_{ij} x_{i'j} \right)^d$$

Em que d é o grau de um polinômio Kernel, generalizando o SVC.

Esta é uma generalização do SVC para um polinômio kernel de grau d . Como exemplificado por James *et al.* (2013), quando o SVC é combinado com um kernel não linear, o resultado é o SVM, um classificador mais flexível e eficaz para diversos problemas. Sua representação pode ser dada por:

$$f(x) = \beta_0 \sum_{i \in S} \alpha_i K(x, x_i)$$

Segundo Zhou, Lai e Yu (2010), o SVM apresenta diversos benefícios quando empregado em problemas de risco de crédito, como é o caso para a presente pesquisa. Para os autores, os principais benefícios são três, a saber:

- i) Requer menos concepções sobre os dados, tais como linearidade e continuidade;
- ii) Possibilita a realização de um mapeamento não linear da estrutura dos dados; e
- iii) É capaz de implementar uma estrutura de minimização de risco (SEM – *Structure Risk Minimization*) a partir de algoritmos que buscam aprender o hiperplano que maximiza as margens.

Uma dificuldade para a aplicação do SVM e do SVC ao problema de análise e concessão de crédito advém de suas características de estimação de parâmetros que não são probabilísticos, mas sim medidas de distâncias (lineares ou não lineares). Assim o SVM e o SVC, para cálculo das probabilidades, dependem de uma aproximação das distribuições de suas estimativas para alguma distribuição de probabilidades.

2.2 Critérios de Decisão em Modelos de Análise de Risco de Crédito

Como dito anteriormente a acurácia é o método mais empregado como métrica comparativa da eficiência dos modelos. Estudos como os de Xiao, Xiao e Wang (2016), Iscoe, *et al.* (2012), Pascual, Martínez e Alamillos (2014), Louzada, Ferreira-Silva e Diniz (2012) e Brown e Mues (2012), empregaram esta métrica como critério de comparação e de consequente seleção de modelos. Como definido por Louzada, Ara e Fernandes (2016) a métrica ACC pode ser calculado como:

$$ACC = \frac{Tp + Tn}{Tp + Fp + Tn + Fn}$$

Em que Tp é o número de classificações a posteriori verdadeiras positivas, Tn é o número de classificações a posteriori verdadeiras negativas, Fp é o número de classificações a posteriori falsas positivas e Fn é o número de classificações a posteriori falsas negativas.

A ACC é uma métrica que calcula o número total de classificações corretas nas predições dos modelos testados. Outras métricas também são observadas em alguns estudos, tais como a Sensitividade, que calcula a taxa de acertos de classificações verdadeiro positivas, a Especificidade, que calcula a taxa de acertos de classificações verdadeiro negativas, dentre outras. Entretanto, todas essas são métricas puramente estatísticas. Como proposta, o presente estudo apresenta o uso do RAROC e do índice de Sharpe como critérios de comparação e consequente decisão para modelos de risco de crédito.

Para Ferson (2013) o RAROC – *Risk Adjusted Return on Capital* ou retorno ajustado ao risco de capital é um indicador que busca medir a rentabilidade de um investimento dado o seu risco, sendo que quanto maior for o retorno em relação ao risco, maior será o desempenho do investimento. De acordo com Ferson (2013), considerando a receita e a despesa de um portfólio de crédito, o RAROC pode ser expresso como:

$$RAROC = \frac{r - e - el + ic}{c}$$

Em que: r são as receitas da carteira, e são as despesas, el é a perda esperada, ic o retorno sobre o capital e c o montante de capital investido na carteira de crédito.

Podem ocorrer variações na formulação do RAROC em função das premissas empregadas (Ferson, 2013). Considerando um processo estocástico, a métrica pode ser definida como: $RAROC = E[r]/EC$, em que $E[r]$ representa a esperança do retorno e EC o capital econômico alocado. Caso seja assumido que o retorno gerado é a riqueza medida pelo EVA® e que o capital econômico alocado é o VaR (*Value at Risk*), o RAROC pode ser medido como sendo: $RAROC = EVA^{\circledast}/VaR$.

Para Ferson (2013), enquanto métrica de avaliação de desempenho financeiro de um investimento, o RAROC considerará a exposição do capital ao risco, sendo expressa como: $RAROC = r/CaR$, em que CaR é o capital em risco (*Capital at Risk*).

Pela proximidade com os problemas tratados em risco de crédito e pela adequação da aplicação e comparação nos modelos de classificação, a pesquisa utilizou esta última formulação do RAROC em sua análise de desempenho financeiro das carteiras, sendo CaR calculado como segue:

$$CaR = VaR_i * CR_i = (-\bar{\varnothing}_i + z\sigma_i) * CR_i$$

Em que: z é o escore padrão da tabela normal para $\alpha=0,05$ unicaldal (aproximadamente 1,65), σ_i é o desvio padrão dos retornos médios das carteiras formadas por cada técnica.

$\bar{\varnothing}_i$ representa a média da variação lognormal dos retornos das carteiras dada por:

$$\bar{\varnothing}_i = \frac{1}{n} \sum \ln \left(\frac{r_{in}}{r_{in-1}} \right)$$

Em que r_{it} representa o retorno da carteira i para cada uma das n simulações realizadas (com $n=100$) e r_{in-1} representa a simulação imediatamente anterior à simulação n .

O índice de Sharpe tem como objetivo oferecer uma medida do retorno de uma carteira de investimento em função do risco da mesma, buscando evidenciar a remuneração do agente frente ao risco ao qual o capital está exposto (Ferson, 2013). Este índice é dado por:

$$IS = \frac{E(R_i - R_f)}{\sigma_{R_i}}$$

Em que: R_i representa o retorno obtido pela carteira de crédito, R_f representa o retorno obtido a partir de uma carteira de referência de mercado tida como livre de risco³ e σ_{R_i} representa o desvio padrão ou volatilidade do retorno da carteira.

Com base nesses indicadores e considerando a carteira de crédito como um investimento da organização, o modelo a ser selecionado deveria ser aquele capaz de oferecer o maior retorno dado o risco.

De uma forma geral, mesmo que se utilizem outras métricas para avaliação da eficiência financeira do investimento em crédito (carteira de crédito), entende-se como adequado e necessário o emprego de métricas de desempenho financeiro para a comparação dos modelos de classificação e análise de risco de crédito. Essa premissa justifica o fato de o estudo não concentrar esforços para a otimização dos algoritmos de classificação, tendo em vista que o objetivo é comparar a informação proveniente das métricas e não a eficiência específica dos modelos por cada métrica.

3 Procedimentos Metodológicos

Para o presente estudo foi utilizada a base de dados *German Credit Dataset*, disponibilizada por Bache e Lichman (2013). As estimações dos parâmetros dos modelos e as classificações foram realizadas com auxílio do software estatístico R.

Esta base traz dados de crédito de mil clientes pessoas físicas, dos quais 30% estão classificados como "em *default*" a priori. Tal base já foi empregada em diversos estudos, tais como os de Ala'raj e Abbod (2016), Harris (2015) e Brown e Mues (2012) e seu amplo uso serve como um *benchmark* para os resultados obtidos pelos modelos. As variáveis presentes na base são apresentadas na Tabela 1.

O estudo preservou o desbalanceamento dos dados, isto é, as amostras de treinamento e de teste foram extraídas do conjunto com 30% dos cliente em *default* e 70% em não *default*. Todos os dados das variáveis empregadas no cálculo das métricas foram extraídas das médias dos testes realizados junto à base dados. Tais médias foram apuradas após 100 (cem) simulações para cada uma das técnicas estudadas, sendo a validação dos resultados realizada por validação cruzada.

Os dados das classificações foram obtidos pela média de 100 repetições dos testes de classificações realizadas com amostras aleatórias com validação cruzada pelo método *k-fold cross validation* com $k=5$, sendo empregados 200 observações para teste e 800 para treinamento. Neste procedimento, a cada simulação o conjunto de dados é particionado (separado) em porções disjuntas de 200 indivíduos ($1000/k$), sendo 800 utilizados para a estimação dos parâmetros do modelo (procedimento conhecido como treinamento) e o restante (200 indivíduos) são utilizados para validação do modelo (procedimento também conhecido como teste). Em *credit scoring* o teste consiste em comparar o valor predito pelo modelo para cada indivíduo do conjunto de dados de teste com os valores reais já presentes no conjunto de dados (classificações a priori).

Tabela 1:
Variáveis Presentes na Base De Dados

Variável	Descrição	Tipo de Variável	nº Categorias
Checking	Status da conta do solicitante em relação ao salário declarado	Ordinal	4,00
Duration	Duração do empréstimo requerido	Contínua	-
history	histórico de créditos anteriores	Ordinal	4,00
purpose	Finalidade do crédito requerido	Nominal	11,00
amount	Montante do crédito requerido	Contínua	-
savings	Economias do cliente	Ordinal	5,00
employ	Tempo no emprego atual	Ordinal	5,00
rates	Juros praticados (%) no empréstimo requerido	Contínua	-
status	Sexo e estado civil	Nominal	5,00
debtors	Outras dívidas	Nominal	3,00
residence	Tempo na residência atual	Contínua	-
property	Propriedades	Nominal	4,00
age	Idade	Contínua	-
other_inst	Outros planos de parcelamento	Nominal	3,00
housing	Tipo de moradia	Nominal	3,00
exist_cr	Nº de créditos concedidos no banco	Contínua	-
job	Profissão	Ordinal	4,00
provider	Número de dependentes	Contínua	-
phone	Telefone próprio	Binária	2,00
foreign	Estrangeiro	Binária	2,00
goodbad	Qualificação do cliente como bom ou mal pagador	Binária	2,00

Fonte: Bache e Lichman (2013)

Os testes geram resultados nos quais o valor predito pode ser o mesmo que o valor a priori ou não. São chamados verdadeiros positivos ou verdadeiros negativos os clientes cujas classificações são “não *default*” ou “*default*” (respectivamente) a priori e as estimadas coincidiram. Da mesma forma, são chamados de falsos positivos e falsos negativos os clientes cujas classificações como “*default*” ou como “não *default*” (respectivamente) a priori e as estimadas não coincidiram.

Pelo método *k-fold* o subconjunto de teste é alterado *k* vezes, sendo que, ao final das *k* interações são calculados os resultados dos acertos (verdadeiros) e dos erros (falsos) encontrados. O objetivo desse procedimento é a obtenção de resultados mais confiáveis sobre a capacidade preditiva do modelo.

Caso se utilizasse o modelo de *credit scoring* como critério único de aprovação de solicitações de crédito, a carteira seria formada pelas solicitações classificadas a priori como positivas (sendo elas falsas ou não). Partindo-se dessa premissa, as métricas RAROC e IS poderiam ser utilizadas como mecanismo de comparação do desempenho de modelos de *credit scoring*, uma vez que os índices poderiam calcular o desempenho financeiro da carteira formada por cada modelo. Esta é a premissa empregada pela pesquisa para comparação financeira dos modelos, isto é, entende-se como viável comparar o desempenho dos modelos com base no desempenho das carteiras que seriam formadas por cada um no caso de sua aplicação.

Na Figura 4 é apresentado o fluxograma dos procedimentos para estimação dos resultados.

Considerando as formulações das métricas RAROC e IS evidenciadas em 2.2, os dados para seu cálculo foram extraídos como exposto na Tabela 2. Neste estudo, foram comparados modelos gerados pelas já citadas técnicas: análise discriminante, regressão logística e redes bayesianas. Cabe ressaltar que o objetivo aqui não é a máxima acurácia do modelo, mas sim sua eficiência financeira.

Figura 4. Fluxograma de procedimentos para estimação

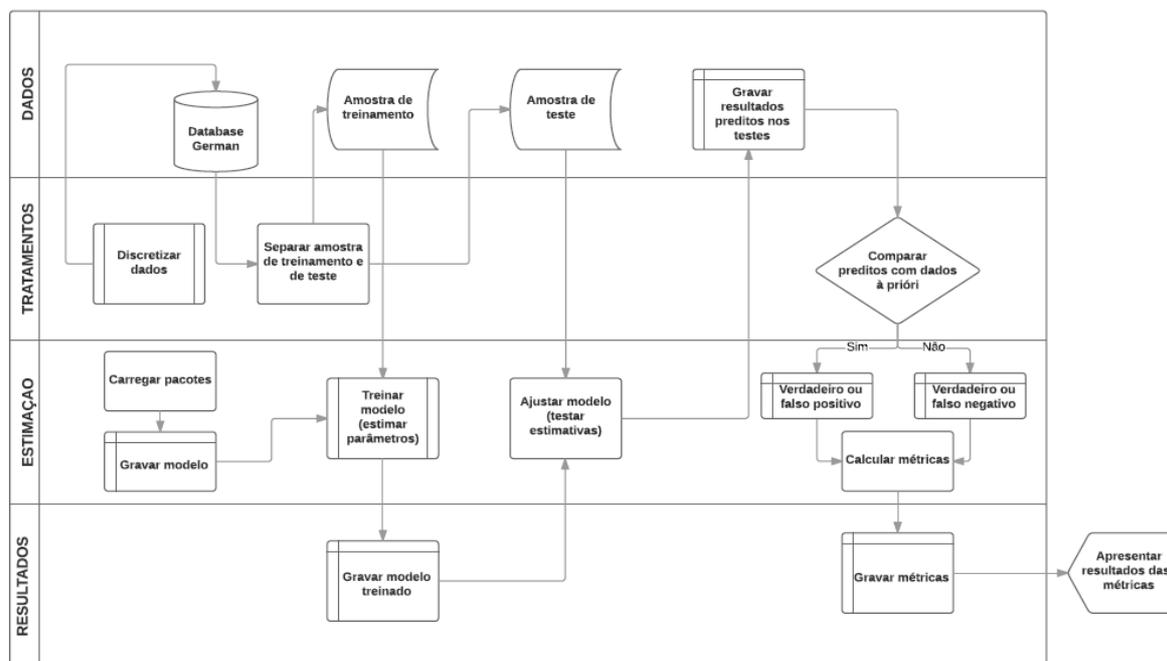


Tabela 2:
Levantamento dos dados das métricas

Métricas	Variáveis	Forma de levantamento dos dados
IS	R_i	Taxa declarada na variável <i>rates</i> para o <i>i</i> -ésimo solicitante multiplicada pela solicitação de cada cliente classificado como não- <i>default</i> .
	R_f	Juros de títulos americanos + Risco-país alemão*
	σ_{R_i}	Desvio padrão médio do retorno da carteira
RAROC	r	Mesmo procedimento adotado em R_i
	CaR	VaR das carteiras simuladas nas $n=100$ classificações realizadas para cada técnica, multiplicado pelo montante emprestado aos clientes classificados como não- <i>default</i> .

Nota: * Para R_f foi empregada como *próxy* a taxa de juros básica dos títulos americanos corrigida pelo risco país alemão por ser uma base de dados alemã. Tais dados foram extraídos de Damodaran (2015).

Os algoritmos são aplicados a cada uma das simulações, sendo os resultados de cada simulação gravados e os resultados apresentados em termos médios ao final da estimação. Empregou-se a curva ROC – *Receiver Operating Characteristic* – como ponte de corte para as classificações. Como destacado por Duan e Shrestha (2011), esse é um método largamente empregado em aplicações de *credit rating* e *credit scoring* para determinação do ponto de corte (*cut-off*), ponto este que delimita a classe a que cada indivíduo pertencerá a posteriori.

A curva ROC é estimada em um intervalo de 0 a 1, considerando as proporções dos erros e acertos de cada classificação para todos os valores de corte, sendo selecionado aquele ponto capaz de promover a maximização das medidas de sensibilidade e especificidade das classificações (Duan & Shrestha, 2011). Para o presente estudo, este valor foi calculado com auxílio do software estatístico R.

Adicionalmente, observou-se que o conjunto de dados traz uma sugestão para cálculo de desempenho financeiro dos modelos por meio da estimativa de custos por erro de classificação. Tal medida também foi utilizada como métrica de desempenho e foi calculada considerando o custo por erro do tipo falso positivo como 1/5 do custo por erro do tipo falso negativo, seguindo as indicações descritas na base de dados.

4 Apresentação e Análise dos Resultados

Seguindo os procedimentos adotados por Chung, Ho e Hsu (2011), este estudo optou por não realizar o balanceamento dos dados, deixando as amostras fiéis ao observado no conjunto de dados original. Segundo os autores, o processo de balanceamento de bancos de dados pode prejudicar a acurácia total de modelos bayesianos. Os autores sugerem ainda que os bancos de dados desbalanceados são mais fiéis à realidade e, desta forma, tais características devem ser respeitadas.

Tabela 3:

Resultados das variáveis utilizadas nas métricas financeiras

Variável	Discriminante	Logística	Naive Bayes	KDB-1	KDB-2	SVC	SVM
ri	2.61%	2.82%	2.83%	2.70%	2.49%	2.87%	2.96%
rf	0.96%	0.96%	0.96%	0.96%	0.96%	0.96%	0.96%
σ_{Ri}	8.41%	4.74%	8.63%	5.27%	6.41%	3.97%	3.62%
<i>VaR</i>	-17.98%	-11.03%	-19.16%	-12.29%	-14.55%	-9.43%	-8.52%
<i>CR</i>	511,108.35	1,190,196.62	450,250.45	1,149,258.34	686,841.90	1,291,530.41	1,444,995.70
<i>CaR</i>	91,908.19	131,338.09	86,272.60	141,250.11	99,969.16	121,795.88	123,120.96

Nota: Sendo *ri* o retorno sobre a carteira, *rf* é a taxa livre de risco e σ_{Ri} a volatilidade média do retorno da carteira (desvio padrão), *VaR* o montante de perda máxima com intervalo de confiança de 95%, *CR* a carteira de crédito e *CaR* o montante de capital em risco (*VaR* multiplicado pela *CR*).

A Tabela 3 traz os valores das variáveis que compõem as métricas financeiras RAROC e IS obtidos a partir da aplicação dos procedimentos metodológicos descritos anteriormente.

O Retorno sobre a carteira (*ri*) foi medido pela soma do ganho com juros de empréstimos aos clientes classificados acertadamente como verdadeiros positivos (bons) e do ganho da recuperação de clientes falsos negativos (ruins) em cada um dos modelos testados. Já para a taxa livre de risco (*rf*) optou-se por considerar a taxa de juros sobre títulos da dívida americana em jan/14, somado ao risco país alemão.

A métrica estatística utilizada, como destacado, foi o ACC e as demais métricas financeiras foram o RAROC e o IS. Os resultados estão disponíveis na Tabela 4.

Tabela 4:

Comparação entre os modelos para cada métrica.

Modelo	ACC (%)	Custo	IS (%)	RAROC (%)
Análise discriminante	65.8	132	19.67	14.5
Regressão Logística	73.4	122.08	39.22	25.5
Naive Bayes	70.24	86.69	21.65	14.8
KDB-1	69.13	173.75	33.00	22.0
KDB-2	55	233.6	23.84	17.1
SVC	76	109.44	48.00	30.4
SVM	86.59	50.46	55.14	34.7

Como é possível observar pelos resultados pelo ACC expressos na Tabela 4, o melhor modelo em acerto seria o modelo SVM, obtendo 86.59% de acertos em média, seguido do modelo SVC, de regressão logística, de rede bayesiana Naive Bayes, de rede KDB-1, de análise discriminante e KDB-2 respectivamente.

Os resultados sobre a métrica ACC corroboram com resultados encontrados em outras pesquisas, tais como: Chung, Ho e Hsu (2011), Duan e Shrestha (2011) e Harris (2015) que trazem o SVM como um algoritmo de classificação com elevada acurácia e flexibilidade de aplicação.

Pela Tabela 4 é possível verificar uma alteração na sequência dos modelos em relação à métrica ACC para o custo por erro. Ainda que os modelos SVM continuem na primeira, a segunda colocação passa a ser do modelo de redes bayesianas Naive Bayes. Em grande parte, o resultado é motivado pela redução da aprovação do Naive Bayes que gerou uma carteira menor que a dos demais modelos, gerando, portanto, um menor custo por erro.

O principal benefício do RAROC e do Índice de Sharpe está associado justamente à necessária ponderação do retorno pelo risco que elimina o efeito do tamanho da carteira. Como abordado por Ferson (2013), quanto maior for o Índice de Sharp mais favorável será a relação risco vs retorno. A mesma lógica é empregada para a interpretação dos resultados do RAROC.

No caso do Índice de Sharpe o modelo com maior indicador é o modelo gerado pelo SVM, seguido dos modelos SVC, regressão logística, KDB-1, KDB-2, Naive Bayes e análise discriminante respectivamente. Nota-se, portanto, que o tamanho da carteira, ponderada pelo risco, gerou a diferença que levou o Naive Bayes da terceira e da segunda colocações nas métricas ACC e Custo respectivamente para a sexta colocação no índice de Sharpe. Essa é a mesma posição que o modelo ocupa no RAROC que conserva exatamente a mesma sequência do Índice de Sharpe, qual seja: SVM, SVC, Regressão Logística, KDB-1, KDB-2, Naive Bayes e análise discriminante.

Chama a atenção, porém, a diferença proporcional entre o Índice de Sharpe e o RAROC. A diferença observada mostra que o índice de Sharpe apresenta, teoricamente, maior variação entre os diversos modelos, sendo mais sensível, possivelmente, pela sua ponderação pelo desvio padrão médio do retorno da carteira.

5 Conclusões e Recomendações

A aprendizagem estatística e a estatística computacional vêm sendo amplamente empregadas para tratamento de problemas de análise e concessão de crédito. Este uso tem se intensificado na medida em que novas e modernas ferramentas computacionais são desenvolvidas. No âmbito acadêmico é vasta a produção científica que busca comparar os diversos possíveis modelos de análise e concessão de crédito, buscando aprimorar o conhecimento acerca do tema.

Como observado no referencial aqui apresentado, os estudos nessa área têm empregado métricas estatísticas de acurácia no processo de comparação dos modelos. Tal uso intensivo sugere que a tomada de decisão sobre a seleção do modelo deve ser baseada na capacidade preditiva dos modelos. Entretanto, partindo-se da premissa de que o desempenho da carteira é o resultado financeiro da carteira e não a capacidade preditiva do modelo, empregar como métrica de avaliação e decisão de seleção de modelos apenas a acurácia não se mostra um processo decisório teoricamente eficiente, dadas as variáveis que influenciam o desempenho da carteira e que não são observadas nos cálculos de acurácia.

Com vistas a demonstrar a necessidade do emprego de métricas financeiras na tomada de decisão de seleção de modelos de *credit scoring*, o presente estudo analisou comparativamente o desempenho estatístico e financeiro de modelos baseados em sete das principais técnicas de aprendizagem estatística utilizadas para tratamento do problema de análise e concessão de crédito. O estudo partiu da premissa de que o processo de decisão sobre um modelo deve considerar a capacidade de geração de resultados financeiros desse modelo. Desta forma, testando-os como um modelo de aprovação automática, as carteiras teriam perfis selecionados pelo próprio modelo.

O estudo considerou a base de dados *German Credit Dataset* para estimar os modelos, bem como as quatro métricas de desempenho selecionadas para o estudo, a saber: acurácia total, custo por erro, retorno ajustado ao risco de capital e índice de Sharpe. As estimações dos modelos foram realizadas com auxílio do software estatístico R e dos pacotes MASS e e1071, sendo os resultados apurados como médias de 100 (cem) simulações validadas pelo método *k-fold cross validation*.

Ainda que o modelo baseado no SVM tenha apresentado os melhores resultados comparativos em todas as métricas, o mesmo não aconteceu para os demais modelos, havendo sensíveis variações nos rankings de classificação dos modelos (do melhor para o pior). Além dessa variação, observou-se também grande variação proporcional para cada indicador estudado. Demonstrou-se que pela métrica estatística ACC, por exemplo, o SVM superou em 17,8% a regressão logística (amplamente empregada para problemas de *credit scoring*). Já pelas métricas financeiras IS - índice de Sharp e RAROC esta diferença foi de 40,6% e 36,1% respectivamente.

Dada a simples interpretação e replicação da regressão logística é possível que os tomadores de decisão sejam induzidos a selecioná-la, entendendo que 17,8% de desempenho preditivo não seria algo tão destrutivo. Entretanto, quando se observa as diferenças de desempenho econômico-financeiro da carteira para a base estudada, esse resultado mostra-se importante.

Assim, estudos voltados para o fornecimento de informações úteis à tomada de decisão sobre seleção de modelos de análise e concessão de crédito como os que se apresentam aqui demonstram ser relevantes, especialmente para o quadro tecnológico atual, no qual observa-se grande avanço na construção e aplicação de ferramentas e modelos para tais problemas, tanto para organizações do setor financeiro (cooperativas de crédito e bancos), quanto para outras organizações que atuam com a concessão de crédito. O presente estudo constata-se particularmente interessante pela observância da necessidade de critérios de desempenho financeiros no processo de seleção. Futuros

estudos, voltados para a inclusão de novas métricas financeiras, bem como em novos bancos de dados mostram-se relevantes para a observação da eficiência do processo de seleção de modelos.

Notas

1. KMV é o nome dado ao modelo de gestão de risco de crédito da empresa de consultoria KMV. Ver em: www.kmv.com
2. Para definições detalhadas sobre hiperplano em classificadores observar James *et al.* (2013, cap. 9).
3. Aqui baseada na taxa de retorno dos títulos do governo dos Estados Unidos somada ao risco país Alemão (tendo em vista que a base de dados é daquele país).

Agradecimentos

Evandro M. S Ribeiro agradece à FUNDACE (Fundação para Pesquisa e Desenvolvimento da Administração, Contabilidade e Economia) pelo suporte financeiro para desenvolvimento da pesquisa sob auxílio N° 0111_1/2015.

Referências

- Ala'Raj, M., & Abbod, M. F. (2016). Classifiers consensus system approach for credit scoring. *Knowledge-Based Systems*, 104(15), 89–105.
- Bache, K., & Lichman, M. (2013). *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Science.
- Bluhm, C., Overbeck, L. & Wagner, C. (2010). *Introduction to credit risk modeling* (2ªed.). London: Chapman & Hall, 386p.
- Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39, 3446–3453.
- Chung, H. Y., Ho, C. H., & Hsu, C. C. (2011). Support vector machines using Bayesian-based approach in the issue of unbalanced classifications. *Expert Systems with Applications*, 38(9), 11447–11452.
- Duan, J. C., & Shrestha, K. (2011). Statistical Credit Rating Methods. *Global Credit Review*, 1, 43-64.
- Febraban, *Relatório Anual*. (2013). Recuperado em 05 de dezembro de 2014, de www.febraban.org.br/.
- Ferson, W. E. (2013). *Investment Performance: A Review and Synthesis*. In. Handbook of the Economics of Finance, 2, part B.
- García, F., Giménez, V., & Guijarro, F. (2013). Credit risk management: A multicriteria approach to assess creditworthiness. *Mathematical and Computer Modelling*, (57), 2009-2015.
- Glantz, M., & Kissell, R. (2014). *Multi-asset risk modeling: techniques for a global economy in an electronic and algorithmic trading era*. San Diego: Academic Press.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis*. (7ª ed.) New Jersey. Pearson Prentice Hall.
- Harris, T. (2015). Credit scoring using the clustered support vector machine. *Expert Systems with Applications*, 42(2), 741–750.
- Iscoe, I., Kreinin, A., Mausser, H., & Romanko, O. (2012). Portfolio credit-risk optimization. *Journal of Banking & Finance*, 36(6), 1604–1615.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with application in R*. New York: Springer.
- Louzada, F., & Ara, A. (2012). Bagging k-dependence probabilistic networks: An alternative powerful fraud detection tool. *Expert Systems with Applications*, 39, 11583–11592.
- Louzada, F., Ferreira-Silva, P. H., & Diniz, C. A. R. (2012). On the impact of disproportional samples in credit scoring models: An application to a Brazilian bank data. *Expert Systems with Applications*, 39(9), 8071–8078.

Louzada, F., Ara, A., & Fernandes, G. B. (In Press). Classification methods applied to credit scoring: Systematic review and overall comparison. *Surveys in Operations Research and Management Science*, (Previous screen in November/2016).

Lu, F. Q., Huang, M.; Ching, W. K., & Siu, T. K. (2013). Credit portfolio management using two-level particle swarm optimization. *Information Sciences*, 237(10), 162-175.

Nagarajan, R., Scutari, M., & Lèbre, S. (2013). *Bayesian Networks in R: with Applications in Systems Biology*. Nova York: Springer.

Pascual, M. B., Martínez, A. M., & Alamillos, A. M. (2014). Redes bayesianas aplicadas a problemas de credit scoring. Una aplicación práctica. *Cuadernos de Economía*, 104(37), 73-86.

Xiao, H., Xiao, Z., & Wang, Y. (2016). Ensemble classification based on supervised clustering for credit scoring. *Applied Soft Computing*, 43, 73-86.

Zhou, L., Lai, K. K., & Yu, L. (2010). Least squares support vector machines ensemble models for credit scoring. *Expert Systems with Applications*, 37(1), 127-133.