

UTILIZANDO APRENDIZADO DE MÁQUINA PARA ESTIMAÇÃO DO *SPREAD* DAS INSTITUIÇÕES FINANCEIRAS NOS EMPRÉSTIMOS DO BNDES

Patrick Franco Alves¹
João Alberto De Negri²
Eric Jardim Cavalcante³

1 INTRODUÇÃO

A essência da atividade bancária consiste na captação de depósitos por meio de poupanças e outras fontes, e na concessão de empréstimos, o que ajuda na alocação de fundos de maneira mais eficiente na economia. A intermediação bancária garante às instituições financeiras uma margem de lucro, também denominada *spread* bancário e comumente definida como receita de juros menos despesas de juros. O princípio da cobrança do *spread* é remunerar a instituição pela sua exposição ao risco de crédito ao emprestar o dinheiro ao cliente. Essa remuneração deve ser superior ao custo de captação e somada às perdas esperadas de inadimplência, aos custos administrativos e outros custos.

Visando ao monitoramento de tendências de mercado e da eficiência dos bancos, a literatura tem se ocupado em entender quais os determinantes do *spread* bancário. *Spreads* bancários muito altos podem refletir um ambiente bancário regulatório inadequado. Uma preocupação é quanto à cobrança de taxas de juros injustas e inacessíveis aos consumidores e empresas (Hawtrey e Liang, 2008). Desta forma, o Banco Nacional de Desenvolvimento Econômico e Social (BNDES) tem manifestado sua preocupação em relação à adequação dos *spreads* cobrados em operações que utilizam recursos subsidiados, sejam elas diretas ou indiretas.⁴ Da mesma maneira, algumas iniciativas do Banco Central do Brasil (BCB) têm procurado incentivar a diminuição dos *spreads* bancários mediante o aumento da competição de mercado (Dantas, Medeiros e Capelletto, 2011).

1. Pesquisador do Programa de Pesquisa para o Desenvolvimento Nacional (PNPD) na Diretoria de Estudos e Políticas Setoriais de Inovação e Infraestrutura (Diset) do Ipea. *E-mail*: <patrick.alves@ipea.gov.br>.

2. Técnico de planejamento e pesquisa na Diset/Ipea. *E-mail*: <joao.denegri@ipea.gov.br>.

3. Pesquisador do PNPD na Diset/Ipea. *E-mail*: <eric.cavalcante@ipea.gov.br>.

4. Disponível em: <<https://bit.ly/3asqCaT>>.

As equações de formação do *spread* no nível macro de Hawtrey e Liang (2008) consideram concentração de mercado, aversão ao risco e tamanho do mercado como os principais determinantes do *spread* bancário:

$$s = \frac{\alpha}{\beta} + \frac{1}{2} R \sigma_I^2 Q \quad (1)$$

Em que:

- s : *spread*;
- $\frac{\alpha}{\beta}$: concentração de mercado;
- R : taxa de aversão ao risco;
- σ_I^2 : variância da taxa de juros da economia; e
- Q : tamanho das transações bancárias.

O aumento da concentração de mercado estaria associado ao aumento do *spread* bancário. O efeito da aversão ao risco (R) sobre o *spread* dependeria também dos movimentos simultâneos na volatilidade da taxa de juros da economia (σ_I^2) e do tamanho do mercado (Q). Entretanto, de forma geral, maior aversão ao risco resultaria em maiores *spreads* bancários.

Subtraindo a concentração de mercado dos dois lados da equação e aplicando o logaritmo natural temos:

$$s - \frac{\alpha}{\beta} = \frac{1}{2} R \sigma_I^2 Q \quad (2)$$

$$\ln\left(s - \frac{\alpha}{\beta}\right) = \ln\left(\frac{1}{2}\right) + \ln(R) + 2 \ln(\sigma_I) + \ln(Q)$$

Considerando a concentração de mercado como um fator fixo, do ponto de vista macroeconômico é possível expressar o *spread* como uma relação linear entre taxa de aversão ao risco, variabilidade da taxa de juros e tamanho do mercado.

$$\ln(s) = \beta_0 + \beta_1 \ln(R) + \beta_2 \ln(\sigma_I) + \beta_3 \ln(Q) \quad (3)$$

A equação (3) é uma relação macroeconômica, e mesmo assim já representa uma relativa simplificação dos determinantes do *spread* bancário. Dias e Ichikava (2011) testam muitos outros determinantes, inclusive os índices de bolsa de valores no mercado americano. Entre todos os indicadores verificados, destaca-se a taxa de inadimplência, que seria capturada pelo índice de aversão ao risco.

Hawtrey e Liang (2008) defendem também que a complexidade no entendimento dos determinantes do *spread* aumenta à medida que nos movemos da representação no nível macroeconômico para o nível mesoeconômico (setorial) e microeconômico (firmas e consumidores).

2 OTIMIZAÇÃO DE HIPERPARÂMETROS

Muitos dos avanços em aprendizado de máquina resultam do aprimoramento e melhores configurações de técnicas já existentes, e não necessariamente de abordagens inéditas (Bergstra *et al.*, 2011). Este é o caso também dos métodos de otimização de hiperparâmetros, que visam extrair o máximo de *performance* de algoritmos já existentes. Um algoritmo muito sofisticado com os hiperparâmetros incorretos pode ter pior desempenho do que um algoritmo simples com hiperparâmetros adequados (Hastie, Tibshirani e Friedman, 2001). A otimização de hiperparâmetros tem o intuito de encontrar os melhores hiperparâmetros para um determinado algoritmo, considerando certo conjunto de dados, eliminando a necessidade de basear-se na consulta prévia ou na tentativa e erro.

Essa otimização tornou-se uma tarefa importante em aprendizado de dados. Os hiperparâmetros não são aprendidos pelo algoritmo durante o treino, e, portanto, precisam ser adequadamente escolhidos previamente. Sem uma metodologia adequada para escolha e definição de tais hiperparâmetros, seria necessário estimar o mesmo algoritmo múltiplas vezes, modificando os hiperparâmetros conforme alguma intuição. Em cada tentativa seria preciso anotar as métricas de precisão, para posteriormente definir qual a melhor combinação de hiperparâmetros. Foi para evitar a ineficiência de tal abordagem que surgiram os métodos de otimização de hiperparâmetros. Entre os métodos existentes, podemos citar otimização bayesiana (BS), *grid search* (GS), *random search* (RS), *halving grid search* (HGS), *halving random search* (HRS), entre outros.

O GS é a abordagem mais simples e mais demorada. Nesta abordagem deve-se estimar todas as combinações possíveis do espaço de hiperparâmetros, armazenando as métricas de precisão em cada interação e finalmente apontando qual a melhor combinação de hiperparâmetros. Em vez de percorrer exaustivamente todas as combinações possíveis dentro do espaço de hiperparâmetros, o RS, por sua vez, realiza o sorteio aleatório de algumas combinações. Assim, é possível diminuir significativamente o tempo necessário para encontrar os hiperparâmetros ótimos, entretanto, sob o risco de não testar alguma combinação de hiperparâmetros ideal. Embora o método GS seja exaustivo, a literatura recente aponta que as chances de encontrar hiperparâmetros ótimos são maiores no RS (Bergstra e Bengio, 2012). Isto porque, na prática, geralmente se restringe o espaço de hiperparâmetros durante o GS.

Ambos os métodos GS e RS desconsideram sequência temporal do processo de otimização. Ao desprezar a história contada pelas rodadas da otimização descartam-se algumas informações que poderiam apontar para uma melhor direção de investigação. Por exemplo, algumas regiões no espaço de hiperparâmetros mereceriam maior investigação, enquanto outras regiões poderiam ter sido abandonadas prematuramente. O método bayesiano e os métodos de *halving* prometem

acelerar o processo de otimização ao fazer uso das informações geradas em cada rodada de interações.

No método bayesiano, o resultado de execuções anteriores ajuda a melhorar as escolhas de hiperparâmetros das próximas interações. Neste método deve-se definir uma função objetivo que auxilie na busca dos hiperparâmetros ideais. Esta função objetivo é atualizada em cada interação utilizando as métricas de previsão e hiperparâmetros da etapa anterior (Brochu, Vlad e Freitas, 2010). Embora prometa grandes melhorias no processo de otimização de hiperparâmetros, as atuais implementações do método bayesiano não permitem a otimização de hiperparâmetros não numéricos. Por exemplo, a função de penalidade (*squared-loss*, *huber*, *epsilon-insensitive*, *squared-epsilon-insensitive*) do algoritmo *stochastic gradient descent* (SGD) e a função de ativação (*identity*, *logistic*, *tanh*, *relu*) no *multilayer perceptron* (MLP) não são características numéricas, não sendo objeto de otimização do método bayesiano. Aparentemente as atuais aplicações do método geralmente se concentram em apenas um hiperparâmetro. A título de exemplo, no algoritmo *multi-task elastic-net* (*multi-task*), otimizaríamos somente o hiperparâmetro alfa, mantendo os demais hiperparâmetros fixos, especialmente os parâmetros categóricos.

Os métodos HGS e HRS não utilizam uma função objetivo. A diminuição do tempo de otimização é atingida ao se restringir o tamanho da amostra nas interações iniciais, descartando definitivamente candidatos a hiperparâmetros ruins nessas amostras menores. O tamanho da amostra de dados aumenta com o amadurecimento dos candidatos a hiperparâmetros, depois que candidatos ruins são definitivamente descartados. Entretanto, se os dados forem excessivamente heterogêneos, como é o caso de informações econômico-financeiras, as amostras mais reduzidas utilizadas nos descartes iniciais dos hiperparâmetros podem não refletir a variabilidade do banco de dados completo. Consequentemente, é possível descartar prematuramente hiperparâmetros que poderiam ser bem-sucedidos utilizando frações maiores da amostra.

Neste trabalho escolhemos os métodos GS e RS, deixando os demais métodos para pesquisa futura. Como nosso interesse é primordialmente a análise das importâncias das características (*features importances*), a utilização desses dois métodos de otimização é suficiente para alcançar nossos objetivos.

3 BASE DE DADOS

A base de dados utilizada para a análise é composta por uma concatenação de três diferentes fontes de informações: Relação Anual de Informações Sociais do Ministério do Trabalho e Emprego (Rais/MTE), Estatísticas Bancárias Mensais (Estbans) do BCB e *spread* das operações indiretas de empréstimos do BNDES proveniente também do BCB.

- Informações de empresas (Rais): pessoal ocupado total, idade média, massa salarial, pessoal ocupado com nível superior, taxa de rotatividade, natureza jurídica, porcentagem de engenheiros e participação de mercado.
- Informações dos bancos (Estbans): conglomerado, segmento (B1, B2, B3, B4, N1) (tabela 1), ativo total, carteira de crédito do banco, passivo circulante, captações de mercado, patrimônio líquido, lucro líquido, número de agências bancárias e número de postos de atendimento.
- Informações das operações (*spread*): valor da operação de crédito, *spread* BNDES, *spread* das instituições financeiras (*target*), taxa de juros, prazo de amortização e prazo de carência.

TABELA 1
Quantidade de operações por segmento bancário das instituições financeiras

Tipo de consolidado bancário	Frequência	%
B1	34.647	58,63
B2	22.218	37,60
B3	229	0,39
B4	1.447	2,45
N1	556	0,94

Fonte: Estban/BCB.
Elaboração dos autores.

As informações da Rais e Estban abrangem mais de dez anos de cobertura. Entretanto, o período utilizado na análise é limitado pelos dados de *spread* do BNDES, cuja cobertura abrange 2016 e 2017. Ainda, os dados de *spread* do BNDES abarcam somente as operações indiretas, as quais possuem uma instituição financeira intermediadora. Apesar de os recursos serem do BNDES, essas instituições atribuem uma avaliação de risco, assumem algum risco de inadimplência na intermediação financeira e cobram seu respectivo *spread* bancário. Por consequência, os dados contêm o *spread* do BNDES e o *spread* das instituições financeiras. Os dados de *spread* bancário do BNDES podem ser concatenados com as informações dos bancos (Estban) por meio do Cadastro Nacional da Pessoa Jurídica (CNPJ) da instituição intermediadora disponível em ambos os bancos de dados.

As variáveis explicativas (*features*) são todas as demais informações descritas em cada conjunto de dados, sendo incluídas também as interações entre essas variáveis. Não fizemos, contudo, interações entre variáveis de conjuntos de dados diferentes. Isto porque nosso objetivo é verificar qual conjunto de informações (firmas, bancos, operações) mais contribui na predição do *spread*. Finalmente, aplicamos logaritmo natural nas variáveis quantitativas e *encoding* de frequência nas variáveis de nominais (segmento bancário, conglomerado e natureza jurídica).

4 ESCOLHA DO ESPAÇO DE HIPERPARÂMETROS

Os efeitos da mudança de hiperparâmetros sobre as métricas de precisão podem depender de diversos fatores, tais como o tamanho dos dados, o número de variáveis explicativas, a natureza da variável dependente ou *target* (binária, contínua, multicategórica), a variabilidade dos dados, entre outros. Nem sempre é claro qual a contribuição de cada hiperparâmetro para evitar subajuste (*underfitting*) ou sobreajuste (*overfitting*) dos dados. Por exemplo, sabe-se que no *extreme gradient boosting* (XGBoost) a profundidade das árvores de decisão (*max-depth*) pode melhorar as métricas de decisão e ao mesmo tempo ocasionar o *overfitting*. O mesmo pode ser dito sobre o número de camadas no algoritmo MLP. Entretanto, uma explicação exaustiva sobre a finalidade de cada parâmetro em cada um dos algoritmos foge do escopo deste capítulo. Em vez disso, definiremos um espaço de hiperparâmetros relativamente generoso conforme apresentado no quadro 1, aplicando os métodos de otimização GS e RS.

QUADRO 1
Espaço de hiperparâmetros dos algoritmos

Algoritmo	RS space	GS space
<i>Multi-task</i>	alpha = [0.005 to 4.755 by 0.25] l1-ratio = [0.01 to 0.25 by 0.1] tolerance = [0.0002 to 0.0982 by 0.002] selection = [cyclic, random]	alpha = [0.005 to 4.755 by 0.60] l1-ratio = [0.01 to 0.25 by 0.3] tolerance = [0.0002 to 0.0982 by 0.01] selection = [cyclic, random]
<i>Least-angle regression (LARS)</i>	alpha = [0.0033 to 2.97 by 0.033] fit-intercept = [True, False] fit-path = [True, False] normalize = [True, False] copy-X = [True, False] positive = [True, False] eps = [10 to 99 by 1]	alpha = [0.0033 to 2.97 by 0.1] fit-intercept = [True, False] fit-path = [True, False] normalize = [True, False] copy-X = [True, False] positive = [True, False] eps = [10 to 99 by 3]
SGD	alpha = [0.8 to 9.8 by 0.625] l1-ratio = [0.05 to 0.24 by 0.01] loss = [squared loss, huber, epsilon insensitive, squared epsilon insensitive] penalty = [l2, l1, elastic-net] epsilon = [0.14 to 2.714 by 0.285] learning-rates = [constant, optimal, invscaling, adaptive] eta0 = [0.0008333 to 0.159 by 0.0083] power-t = [0.10 to 0.483 by 0.0167] early_stopping = [False, True]	alpha = [0.8 to 9.8 by 2] l1-ratio = [0.05 to 0.24 by 0.05] loss = [squared loss, huber, epsilon insensitive, squared epsilon insensitive] penalty = [l2, l1, elastic-net] epsilon = [0.14 to 2.714 by 0.7] learning-rates = [constant, optimal, invscaling, adaptive] eta0 = [0.0008333 to 0.159 by 0.01] power-t = [0.10 to 0.483 by 0.05] early_stopping = [False, True]
<i>Least squares with l2 regularization (ridge)</i>	alpha = [0.016 to 1.96 by 0.01] fit-intercept = [True, False] normalize = [True, False] copy-X = [True, False] solver= [auto, svd, cholesky, lsqr, sparse_cg] tolerance = [1e-05 to 0.01 by 0.0001]	alpha = [0.016 to 1.96 by 0.05] fit-intercept = [True, False] normalize = [True, False] copy-X = [True, False] solver= [auto, svd, cholesky, lsqr, sparse_cg] tolerance = [1e-05 to 0.01 by 0.002]

(Continua)

(Continuação)

Algoritmo	RS space	GS space
<i>Regression with l1 and l2 regularizer (elastic-net)</i>	alpha = [0.002 to 0.992 by 0.01] l1-ratio = [0.04 to 2500 by 0.04] fit-intercept = [True, False] normalize = [True, False] copy-X = [True, False] precompute = [True, False] warm-start = [True, False] positive = [True, False] tolerance = [0.001 to 0.999 by 0.001] selection = [cyclic, random]	alpha = [0.002 to 0.992 by 0.05] l1-ratio = [0.04 to 2500 by 1] fit-intercept = [True, False] normalize = [True, False] copy-X = [True, False] precompute = [True, False] warm-start = [True, False] positive = [True, False] tolerance = [0.001 to 0.999 by 0.005] selection = [cyclic, random]
<i>Bayesian ridge (Bayes)</i>	alpha-1 = [0.02 to 20 by 0.02] alpha-2 = [0.02 to 20 by 0.02] lambda-1 = [0.02 to 20 by 0.02] lambda-2 = [0.02 to 20 by 0.02] compute-score = [False, True] copy-X = [False, True] fit-intercept = [False, True] normalize = [False, True] tolerance = [1e-07 to 1e-02 by 1e-07]	alpha-1 = [0.02 to 20 by 0.06] alpha-2 = [0.02 to 20 by 0.06] lambda-1 = [0.02 to 20 by 0.06] lambda-2 = [0.02 to 20 by 0.06] compute-score = [False, True] copy-X = [False, True] fit-intercept = [False, True] normalize = [False, True] tolerance = [1e-07 to 1e-02 by 1e-03]
<i>Least absolute shrinkage selection operator (LASSO)</i>	alpha = [0.001 to 0.095 by 0.005] fit-intercept = [True, False] copy-X = [True, False] normalize = [True, False] precompute = [True, False] positive = [True, False] selection = [cyclic, random] tolerance = [0.0001 to 0.01 by 1e-05]	alpha = [0.001 to 0.095 by 0.01] fit-intercept = [True, False] copy-X = [True, False] normalize = [True, False] precompute = [True, False] positive = [True, False] selection = [cyclic, random] tolerance = [0.0001 to 0.01 by 1e-05]
<i>K-nearest neighbor (KNN)</i>	n-neighbors = [4 to 9 by 1] weights = [uniform, distance] algorithm = [auto, ball-tree, kd-tree, brute] leaf-size = [5 to 34 by 1] p = [2 to 8 by 1]	n-neighbors = [4 to 9 by 1] weights = [uniform, distance] algorithm = [auto] leaf-size = [5 to 34 by 1] p = [2 to 8 by 1]

Elaboração dos autores.

O quadro 1 mostra o espaço de hiperparâmetros para cada um dos algoritmos. É possível observar que o espaço de hiperparâmetros do método GS é mais restrito, o que visa agilizar a otimização. De forma geral, isto foi realizado mantendo-se igual o espaço das funções de ativação para os métodos GS e RS. Já para os parâmetros de tolerância e regularização, usamos a mesma amplitude de espaço do RS, mas com uma distância maior entre os valores (*step*) para o GS.

5 RESULTADOS

Para comparação dos algoritmos vamos utilizar as métricas de acurácia: escore da variância explicada (EVS), *R*-quadrado (R^2) e erro médio absoluto. Além disso, vamos reportar o tempo de estimação necessário. As métricas de acurácia são dadas pelas expressões:

$$EVS = 1 - \frac{VAR(y_i - \hat{y}_i)}{VAR(y_i)} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$$MAE = 1 - \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

A métrica *mean absolute error* (MAE) captura os resíduos gerados pelo modelo, enquanto o R^2 representa a proporção da variância na variável *target* explicada pelo modelo. Por sua vez, o EVS mensura a proporção da variância explicada pelo modelo. Esta é uma métrica interessante para complementar as demais, pois não utiliza o valor médio do *target* (\hat{y}), não sendo influenciada por valores extremos.

A tabela 2 mostra as métricas EVS, R^2 e MAE e o tempo de treinamento para os métodos de otimização RS e GS. Conforme esperado, os tempos de treinamento utilizando otimização exaustiva (GS) são sempre maiores do que quando utilizamos RS. Mesmo quando a métrica EVS é maior para o método de otimização GS, ainda assim é possível argumentar que o ganho de precisão não compensa o tempo de treinamento. Para o método de otimização RS, os algoritmos XGBoost, LASSO, *multi-task*, *Bayes*, *ridge* e LARS apresentaram resultados bem satisfatórios. Particularmente a otimização de hiperparâmetros do algoritmo *ridge* produziu as melhores métricas de precisão EVS e R^2 , enquanto o LARS produziu a melhor estatística MAE.

TABELA 2
Métricas EVS, R^2 , MAE e tempo de treinamento

Algoritmo	RS				GS			
	EVS	R^2	MAE	Time	EVS	R^2	MAE	Time
XGBoost	0.9958	0.9958	0.1197	1h5min5s	0.9482	0.9482	0.0152	8h9min36s
LASSO	0.9076	0.9035	0.0332	39,8s	0.9540	0.9540	0.0915	25min44s
<i>Multi-task</i>	0.9913	0.9913	0.0868	1min3s	0.9607	0.9606	0.0802	36min13s
<i>Bayes</i>	0.9917	0.9917	0.0183	2min38s	0.9559	0.9559	0.0847	4h2min4s
<i>Elastic-net</i>	0.6567	0.6556	0.0745	15,3s	0.5572	0.5701	0.0750	6h47min1s
<i>Ridge</i>	0.9917	0.9917	0.2164	7,25s	0.9538	0.9538	0.0192	1min30s
LARS	0.9909	0.9909	0.2269	49,8s	0.1247	0.1247	0.0262	2h9min2s
SGD	0.3301	0.3301	0.0983	1min55s	0.4832	0.4498	0.0876	2h4min41s
KNN	0.7552	0.7210	0.0193	3min35s	0.9510	0.9507	0.0951	22min46s

Elaboração dos autores.

Os quadros 2, 3 e 4 apresentam as quinze principais importâncias das características para os algoritmos XGBoost, *multi-task* e LARS. Nesses quadros procuramos identificar as características que melhor explicam a predição do *spread* bancário das instituições financeiras. A análise procura ressaltar se as características mais importantes dizem respeito aos dados da operação bancária (dados de *spread* do

BNDES), dos bancos intermediadores (Estban) ou das empresas solicitantes dos financiamentos (Rais).

QUADRO 2
Principais importâncias das características para XGBoost

Fonte	Variável explicativa	Importância das características
<i>Spread</i> BNDES	Interação entre <i>spread</i> BNDES e juros	0,5353
<i>Spread</i> BNDES	Taxa de juros	0,3443
<i>Spread</i> BNDES	Interação entre valor da operação e juros	0,0859
<i>Spread</i> BNDES	<i>Spread</i> BNDES	0,0019
Estban	Carteira ativa de crédito	0,0019
<i>Spread</i> BNDES	Interação entre prazo de carência e juros	0,0017
Estban	Segmento bancário B2	0,0015
Estban	Número de postos de atendimento	0,0012
Estban	<i>Encoding</i> do conglomerado banco	0,0012
Estban	Interação entre captações de mercado e ativo total	0,0010
Estban	Captações de mercado	0,0009
Estban	Ativo total	0,0009
Estban	Interação entre ativo total e passivo	0,0009
Estban	Carteira de crédito do banco	0,0009
Rais	Tamanho da firma	0,0008

Elaboração dos autores.

QUADRO 3
Principais importâncias das características para *multi-task*

Fonte	Variável explicativa	Importância das características
<i>Spread</i> BNDES	Taxa de juros	0,219
<i>Spread</i> BNDES	<i>Spread</i> BNDES	0,145
Estban	Segmento bancário B3	0,129
Estban	Interação entre carteira de crédito e ativo total	0,048
<i>Spread</i> BNDES	Prazo de amortização	0,045
<i>Spread</i> BNDES	Interação entre <i>spread</i> BNDES e juros	0,041
<i>Spread</i> BNDES	Prazo de carência	0,031
Estban	Carteira de crédito do banco	0,031
Estban	Interação entre carteira de crédito e passivo do banco	0,030
<i>Spread</i> BNDES	Interação entre prazo de carência e juros	0,029
Estban	Interação entre patrimônio de mercado e ativo total	0,024
Estban	Segmento bancário B2	0,024
<i>Spread</i> BNDES	Interação entre prazo de carência e carência	0,022
<i>Spread</i> BNDES	Interação entre valor da operação e <i>spread</i>	0,016
Estban	Passivo total do banco	0,016

Elaboração dos autores.

QUADRO 4
Principais importâncias das características para LARS

Fonte	Variável explicativa	Importância das características
<i>Spread</i> BNDES	Interação entre <i>spread</i> BNDES e juros	0,776
<i>Spread</i> BNDES	Taxa de juros	0,177
<i>Spread</i> BNDES	Interação entre valor da operação e juros	0,019
<i>Spread</i> BNDES	Interação entre prazo de amortização e taxa de juros	0,003
Rais	Interação entre massa salarial e pessoal ocupado de nível superior	0,002
Estban	<i>Encoding</i> do conglomerado banco	0,000
Rais	<i>Encoding</i> da natureza jurídica da firma	0,000
<i>Spread</i> BNDES	Interação entre <i>spread</i> BNDES e prazo de amortização	0,000
<i>Spread</i> BNDES	Interação entre valor da operação e amortização	0,000
<i>Spread</i> BNDES	Interação entre juros e prazo de carência	0,000
Estban	Número de postos de atendimento	0,000
Estban	Número de agências bancárias	0,000
Rais	Pessoal ocupado de nível superior	0,000
<i>Spread</i> BNDES	Interação entre <i>spread</i> BNDES e prazo de carência	0,000
Estban	Lucro líquido do banco	0,000

Elaboração dos autores.

Para o algoritmo XGBoost, apenas o tamanho da firma encontra-se entre as informações mais relevantes para explicar o *spread*. A interação entre taxa de juros e *spread* BNDES destaca-se como sendo a característica mais importante. Várias características da instituição intermediadora podem ser vistas entre as quinze principais importâncias das características.

Para o algoritmo *multi-task*, não encontramos nenhuma característica das empresas entre as informações mais relevantes para explicar o *spread*. A taxa de juros destaca-se como sendo a característica mais importante, seguida pelo *spread* BNDES e o segmento bancário a que pertence a instituição intermediadora. Várias características da instituição intermediadora podem ser vistas entre as quinze principais importâncias das características. Este resultado indica que a política de precificação da instituição financeira envolvida pode ser tão relevante quanto as características das firmas para determinar o *spread* bancário.

Para o algoritmo LARS, após a quinta característica mais importante, todas as demais são iguais a zero. A interação entre o *spread* BNDES e a taxa de juros destaca-se como sendo a característica mais importante para determinação do *spread* bancário das instituições financeiras, seguida da taxa de juros e da interação desta com o valor da operação.

6 CONCLUSÃO

É importante que as instituições financeiras possuam uma rentabilidade compatível com a própria sustentabilidade, o que é exercido por meio da cobrança de *spreads* bancários. Entretanto, recentemente têm surgido questionamentos sobre os elevados lucros alcançados pelos bancos e quanto à cobrança de taxas de juros adequadas. Ademais, *spreads* bancários excessivamente elevados podem indicar um ambiente bancário regulatório inadequado. Neste contexto, este capítulo utilizou o aprendizado de máquina para prever o *spread* bancário das instituições financeiras intermediadoras dos empréstimos do BNDES. Para isso, fizemos uso da métrica de importâncias das características (*features importances*) com o intuito de identificar os principais fatores determinantes desse *spread*.

Juntamente com as características das instituições financeiras, as condições da própria operação, tais como prazo de carência, valor do financiamento e taxa de juros, são os fatores que mais contribuem na determinação do *spread* bancário. Embora a taxa de juros reflita a avaliação de risco da operação, e por consequência as características de contábil das empresas, não podemos descartar que as características das instituições financeiras intermediadoras são fortes preditoras do *spread*. Isto pode indicar que as estratégias negociais, as metodologias de precificação e outras políticas dos bancos são fatores importantes na determinação do *spread* bancário. A força das características das instituições na determinação do *spread* compete com as características das operações, e pode até superar as características das empresas na predição do *spread*.

REFERÊNCIAS

- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, v. 13, n. 10, p. 281-305, 2012. Disponível em: <<https://www.jmlr.org/papers/v13/bergstra12a.html>>.
- BERGSTRA, J. *et al.* Algorithms for hyper-parameter optimization. In: ANNUAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 25., 2011, Granada. **Anais...** Granada: Nips, 2011.
- BROCHU, E.; VLAD, C.; FREITAS, N. **A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning.** [s.l.]: Cornell University, 2010.
- DANTAS, J. A.; MEDEIROS, O.; CAPELLETTO, L. **Determinantes do *spread* bancário *ex post* no mercado brasileiro.** Brasília: BCB, 2011. (Trabalhos para Discussão, n. 242).
- DIAS, V.; ICHIKAVA, E. Uma análise empírica da relação entre *spread* e risco. **Revista do BNDES**, n. 36, 2011.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning**: data mining, inference and prediction. 2nd ed. Stanford: Springer, 2001. (Springer Series in Statistics).

HAWTREY, K.; LIANG, H. Bank interest margins in OECD countries. **The North American Journal of Economics and Finance**, v. 19, n. 3, p. 249-260, 2008.