CONGRESSO INTERNACIONAL DE DIREITO E INTELIGÊNCIA ARTIFICIAL

OS DIREITOS HUMANOS NA ERA TECNOLÓGICA III

FAUSTO SANTOS DE MORAIS

VALTER MOURA DO CARMO

O81
Os direitos humanos na era tecnológica III [Recurso eletrônico on-line] organização Congresso Internacional de Direito e Inteligência Artificial: Skema Business School – Belo Horizonte;

Coordenadores: Valter Moura do Carmo, Felipe Calderón-Valencia e Fausto Santos de Morais – Belo Horizonte: Skema Business School, 2020.

Inclui bibliografia
ISBN: 978-65-5648-099-2
Modo de acesso: www.conpedi.org.br em publicações
Tema: Desafios da adoção da inteligência artificial no campo jurídico.
1. Direito. 2. Inteligência Artificial. 3. Tecnologia. I. Congresso Internacional de Direito e Inteligência Artificial (1:2020 : Belo Horizonte, MG).

CDU: 34

CONGRESSO INTERNACIONAL DE DIREITO E INTELIGÊNCIA ARTIFICIAL

OS DIREITOS HUMANOS NA ERA TECNOLÓGICA III

Apresentação

É com enorme alegria que a SKEMA Business School e o CONPEDI – Conselho Nacional de Pesquisa e Pós-graduação em Direito apresentam à comunidade científica os 14 livros produzidos a partir dos Grupos de Trabalho do I Congresso Internacional de Direito e Inteligência Artificial. As discussões ocorreram em ambiente virtual ao longo dos dias 02 e 03 de julho de 2020, dentro da programação que contou com grandes nomes nacionais e internacionais da área, além de 480 pesquisadoras e pesquisadores inscritos no total. Estes livros compõem o produto final deste que já nasce como o maior evento científico de Direito e da Tecnologia do Brasil.

Trata-se de coletânea composta pelos 236 trabalhos aprovados e que atingiram nota mínima de aprovação, sendo que também foram submetidos ao processo denominado double blind peer review (dupla avaliação cega por pares) dentro da plataforma PublicaDireito, que é mantida pelo CONPEDI. Os quatro Grupos de Trabalho originais, diante da grande demanda, se transformaram em 14 e contaram com a participação de pesquisadores de 17 Estados da federação brasileira. São cerca de 1.500 páginas de produção científica relacionadas ao que há de mais novo e relevante em termos de discussão acadêmica sobre os temas Direitos Humanos na era tecnológica, inteligência artificial e tecnologias aplicadas ao Direito, governança sustentável e formas tecnológicas de solução de conflitos.

Os referidos Grupos de Trabalho contaram, ainda, com a contribuição de 41 proeminentes professoras e professores ligados a renomadas instituições de ensino superior do país, os quais indicaram os caminhos para o aperfeiçoamento dos trabalhos dos autores. Cada livro desta coletânea foi organizado, preparado e assinado pelos professores que coordenaram cada grupo. Sem dúvida, houve uma troca intensa de saberes e a produção de conhecimento de alto nível foi, certamente, o grande legado do evento.

Neste norte, a coletânea que ora torna-se pública é de inegável valor científico. Pretende-se, com esta publicação, contribuir com a ciência jurídica e fomentar o aprofundamento da relação entre a graduação e a pós-graduação, seguindo as diretrizes oficiais. Fomentou-se, ainda, a formação de novos pesquisadores na seara interdisciplinar entre o Direito e os vários

campos da tecnologia, notadamente o da ciência da informação, haja vista o expressivo número de graduandos que participaram efetivamente, com o devido protagonismo, das atividades.

A SKEMA Business School é entidade francesa sem fins lucrativos, com estrutura multicampi em cinco países de continentes diferentes (França, EUA, China, Brasil e África do Sul) e com três importantes acreditações internacionais (AMBA, EQUIS e AACSB), que demonstram sua vocação para ensino e pesquisa de excelência no universo da economia do conhecimento. A SKEMA, cujo nome é um acrônimo significa School of Knowledge Economy and Management, acredita, mais do que nunca, que um mundo digital necessita de uma abordagem transdisciplinar.

Agradecemos a participação de todos neste grandioso evento e convidamos a comunidade científica a conhecer nossos projetos no campo do Direito e da tecnologia. Já está em funcionamento o projeto Nanodegrees, um conjunto de cursos práticos e avançados, de curta duração, acessíveis aos estudantes tanto de graduação, quanto de pós-graduação. Até 2021, será lançada a pioneira pós-graduação lato sensu de Direito e Inteligência Artificial, com destacados professores da área.

Agradecemos ainda a todas as pesquisadoras e pesquisadores pela inestimável contribuição e desejamos a todos uma ótima e proveitosa leitura!

Belo Horizonte-MG, 07 de agosto de 2020.

Profª. Drª. Geneviève Daniele Lucienne Dutrait Poulingue

Reitora – SKEMA Business School - Campus Belo Horizonte

Prof. Dr. Edgar Gastón Jacobs

Coordenador Acadêmico da Pós-graudação de Direito e Inteligência Artificial da SKEMA Business School

DECISÕES AUTOMATIZADAS E A AMPLITUDE DO DIREITO À EXPLICAÇÃO NA LGPD: UM ESTUDO DE CASO

AUTOMATED DECISION-MAKING AND THE RIGHT TO AN EXPLANATION IN THE BRAZILIAN GENERAL DATA PROTECTION LAW: A CASE STUDY

Bruno Alberto Soares Oliveira
Yuri Alexandre dos Santos

Resumo

A Lei Geral de Proteção de Dados confere aos titulares de dados pessoais afetados por decisões automatizadas o direito de receber informações claras e adequadas relacionadas ao processo decisório. A lei não estabelece o alcance deste direito à explicação, de modo que este trabalho pretende iniciar uma análise acerca de sua amplitude. O assunto demanda conhecimentos técnicos, de modo que a investigação parte de um estudo de caso que considera dois algoritmos de aprendizado de máquina. Argumenta-se que o grau de arbitrariedade conferido ao agente de tratamento de dados no processo decisório pode impactar a amplitude do direito à explicação.

Palavras-chave: Decisões automatizadas, Direito à explicação, Lei geral de proteção de dados pessoais

Abstract/Resumen/Résumé

Brazilian Data Protection Law stablishes that data subjects who are affected by automated decision-making have a right to receive clear and adequate information regarding the decision-making process. Considering that the law does not stablish the amplitude of such a right, this work intends to analyze it. Given that the matter demands technical knowledge, the investigation is based in a case-study of two different machine learning algorithms. We argue that the degree to which a data processor can take arbitrary decisions may impact the amplitude of the right to an explanation.

Keywords/Palabras-claves/Mots-clés: Automated decision-making, Right to an explanation, Brazilian general data protection law

1 Introdução

Embora a Lei Geral de Proteção de Dados Pessoais (LGPD) (BRASIL, 2018) não tenha expressamente conferido à pessoa natural cujos dados pessoais sejam objeto de tratamento (designada, pela norma, como *titulares*) um direito à explicação, a existência, na lei, de tal direito é evidente e decorre de suas disposições quanto ao princípio da transparência em sede de tratamento de dados pessoais, mencionado no art. 6º, inciso VI, daquela lei, segundo o qual os titulares têm direito a receber "informações claras, precisas e facilmente acessíveis sobre a realização do tratamento" de dados pessoais, e daquelas que disciplinam o direito do titular à revisão de decisões automatizadas, contidas no art. 20, *caput* e § 1º, daquele diploma legal (MONTEIRO, 2018).

No entanto, se a existência do direito é clara, o mesmo não se pode dizer de seu alcance. Isso porque as disposições legais que o embasam são amplas e vagas, não havendo qualquer previsão do que possa ser considerado uma *informação clara e adequada* nem de qual a abrangência dos *critérios* e dos *procedimentos* que devem ser informados aos titulares. Além disso, a lei determina que o direito à explicação não compreende informações protegidas por segredos comercial ou industrial, sem dar pistas da abrangência de tal limitação.

De forma ampla, tido como uma emanação do princípio da transparência, o direito à explicação pode abranger qualquer situação na qual um indivíduo pretenda acessar informações sobre operações de tratamento de dados pessoais que lhe digam respeito (MONTEIRO, 2018). Para os fins deste estudo, porém, apenas uma face deste direito será estudada: aquela que o contextualiza no âmbito das decisões automatizadas passíveis de revisão. Nesta linha, em um sentido mais estrito, direito à explicação é o direito correlato, do ponto de vista do titular, à obrigação imposta pelo art. 20, § 1º, da LGPD à pessoa física ou jurídica que tem o controle decisório sobre a atividade de tratamento de dados pessoais (o controlador, seguindo a conceituação legal) e que, ao fazê-lo, submete o titular a algum tipo de decisão automatizada. Ou seja, *direito à explicação* é *o direito que o titular tem de receber, do controlador, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, excluídas as informações cujo acesso possa violar os segredos comercial e industrial do controlador*.

Neste sentido, o direito à explicação é intrinsecamente ligado ao direito à revisão de decisões automatizadas. Aliás, não há dúvidas de que o exercício efetivo do segundo depende do primeiro, tendo em vista que só se pode questionar, com propriedade, uma decisão se se tem conhecimento acerca dos fatores que levaram a ela, sendo que o grau de arbitrariedade do processo decisório pode ser maior ou menor. O direito à explicação é, portanto, essencial para o exercício do contraditório, direito fundamental que deve ser observado também nas relações privadas.

Assim, este estudo não abordará todas as situações nas quais se pode falar em um direito à explicação, mas apenas aquelas nas quais o titular é submetido aos efeitos de uma

decisão automatizada que poderá ser revista. Além disso, não tratará de todos os tipos de decisões automatizadas, mas apenas aquelas tomadas por algoritmos que se valem de técnicas de aprendizado de máquina na modalidade de aprendizagem supervisionada.

O presente estudo investiga o alcance que o direito à explicação pode ter do ponto de vista técnico. Assim, pretende-se discutir o que pode ser explicado, ou seja, quais informações podem ser dadas a um usuário. Além disso, pretende-se investigar se as informações que podem ser dadas são úteis para que o titular efetivamente as compreenda e para que possa exercer seu direito à revisão caso pretenda alterar a decisão tomada. Em síntese, a principal pergunta que a pesquisa pretende responder é: qual a abrangência do direito à explicação em se tratando de decisões automatizadas tomadas por algoritmos de aprendizagem supervisionada?

Para tentar responder à pergunta, o presente estudo se baseia em uma breve discussão sobre a tensão entre a acurácia e a precisão em algoritmos de aprendizagem supervisionada, bem como em um estudo de caso que busca evidenciar alguns dos problemas técnicos envolvidos no tratamento jurídico do problema proposto.

2 Precisão x Interpretabilidade

Atualmente, algoritmos de aprendizado de máquina vêm sendo utilizados com sucesso em várias aplicações, conseguindo simular o comportamento humano e obtendo resultados extremamente satisfatórios em determinadas tarefas.

Para responder às perguntas propostas, é necessário considerar que, no desenvolvimento de um modelo algorítmico para tomada de decisões, dois aspectos importantes devem ser considerados, a *acurácia* e a *interpretabilidade*. Ambos são, geralmente, conflitantes entre si (MORI; UCHIHIRA, 2019).

A acurácia (tomada como sinônimo de precisão neste texto) pode ser entendida como a capacidade que um algoritmo de aprendizado de máquina possui de concluir com êxito uma dada tarefa, enquanto a interpretabilidade diz respeito à possibilidade de que o algoritmo possa ser compreendido pelos seres humanos (JOHANSSON et al., 2011). Estes conceitos, aqui, são adotados do ponto de vista técnico, computacional, de modo que, como se verá adiante, o fato de um sistema ser interpretável não implica, necessariamente, em atendimento ao direito à explicação para fins jurídicos.

Softwares de aprendizado de máquina que possuem alta precisão são comumente complexos e difíceis de interpretar, motivo pelo qual, em muitos casos, recebem o nome de "sistemas de caixa-preta". Já algoritmos interpretáveis, por outro lado, podem não alcançar uma precisão que justifique seu uso. A Figura 1 apresenta um gráfico que relaciona precisão e interpretabilidade para diferentes métodos de aprendizado de máquina.

A título de exemplo, "sistemas de caixa-preta" que se utilizam de Redes Neurais Artificiais (RNAs) com frequência alcançam resultados com alto grau de assertividade. Por outro lado, sistemas que utilizam aprendizado por Regressão Logística podem ter desempenho muito

menor, mas são, em geral, altamente interpretáveis.

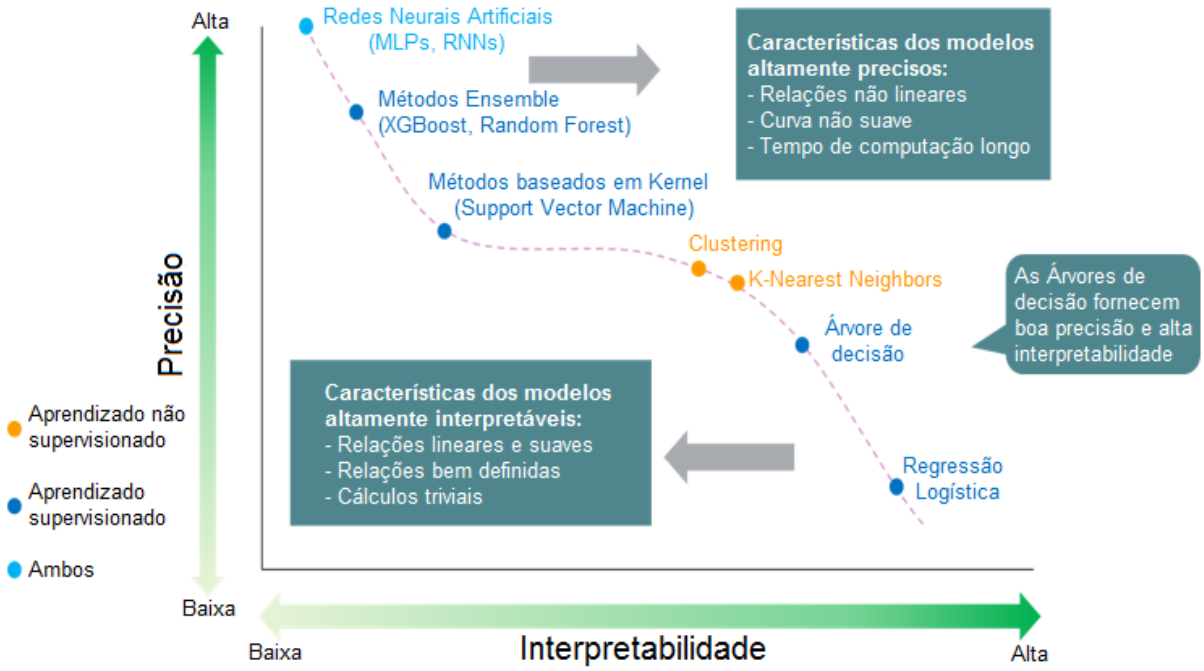

Figura 1: Precisão x Interpretabilidade.

3 Estudo de Caso

3.1 Conjunto de dados

O assunto aqui abordado será problematizado com base em um sistema de análise de risco de crédito, que poderia ser utilizado por uma instituição financeira fictícia para decidir se um determinado cliente que pretenda receber crédito o receberá ou não.

A utilização de algoritmos de inteligência artificial para a análise de risco de crédito possui algumas vantagens, do ponto de vista da instituição que o utiliza, na medida em que lhes permite prever o comportamento de seus clientes, de modo a antever uma eventual inadimplência. Hoje, técnicas avançadas de análise de dados permitem que as empresas analisem o nível de risco de crédito para novos clientes com poucas informações no banco de dados.

O banco de dados utilizado para a análise, o *Statlog (German Credit Data) Data Set* (HOFMANN, 1994), possui informações sobre pessoas que conseguiram, ou não, receber crédito por um banco. No banco de dados, cada indivíduo é classificado como um bom ou mau pagador, dependendo de suas características. O *dataset* possui informações sobre mil pessoas, e vinte características para cada uma delas, além do rótulo “bom” ou “ruim”, que representa a predição que o algoritmo deverá encontrar.

Algumas informações contidas no banco de dados sobre os clientes são: idade, sexo, trabalho (não qualificado, qualificado, altamente qualificado), emprego atual, renda *per capita*, habitação (própria, alugada ou mora de favor), valor da conta corrente atual, valor da conta

poupança atual, valor do pedido do crédito, patrimônio, duração do empréstimo, finalidade do pedido, entre outras.

3.2 Material e Métodos

De uma forma geral, o aprendizado de máquina supervisionado propõe duas tarefas: regressão e classificação. A regressão prediz um valor contínuo (ex.: temperatura, preço etc) e a classificação prediz um rótulo, como no atual estudo de caso, em que o cliente é rotulado como bom ou mau pagador. Serão abordados dois algoritmos de aprendizado de máquina: a Regressão Logística (precisão baixa e interpretabilidade alta) e as Redes Neurais Artificiais (precisão alta e interpretabilidade baixa). Ambos foram implementados utilizando a linguagem de programação *Python*.

Na análise exploratória dos dados (a primeira tarefa de um cientista de dados, na qual busca-se entender o textitdataset e suas informações), constatou-se a existência de setecentas pessoas rotuladas como boas pagadoras e trezentas rotuladas como pagadoras ruins. Tal constatação evidencia um desbalanceamento nos dados, fator que influencia diretamente a precisão do algoritmo, já que poderá implicar em uma maior tendência de novos clientes serem rotulados como pertencentes à classe majoritária (bom pagador).

Sendo assim, aplicou-se uma técnica de balanceamento de dados denominada *SMOTE* (CHAWLA et al., 2002) que, em poucas palavras, cria novos clientes artificiais para a classe minoritária (pagador ruim), de modo a equipará-la, em número de clientes, à classe majoritária. Posteriormente, o conjunto de dados foi dividido aleatoriamente em dois subconjuntos, um de treinamento e outro de teste. Os dados contidos no subconjunto de treinamento são utilizados para encontrar os melhores valores dos parâmetros do algoritmo, enquanto os dados do subconjunto de teste são utilizados para verificar a precisão do método implementado.

3.3 Resultados

A Tabela 1 apresenta os resultados obtidos pela Regressão Logística e pela Rede Neural Artificial para os clientes que pertencem ao subconjunto de testes. Ambos foram implementados utilizando a biblioteca *Scikit-learn* (PEDREGOSA et al., 2011).

Tabela 1: Resultados de classificação obtidos.

	Regressão Logística		**Rede Neural Artificial**	
Precisão	70%		75%	
Matriz de confusão	Bom	Ruim	Bom	Ruim
Bom	160	54	171	43
Ruim	34	52	30	56

Como pode ser observado na matriz de confusão da Tabela 1, de duzentos e catorze

clientes rotulados como bons pagadores no conjunto de teste, o algoritmo de Regressão Logística classificou, corretamente, cento e sessenta, errando quanto aos demais. Ou seja: cinquenta e quatro clientes que eram bons pagadores foram rotulados, pelo algoritmo, como pagadores ruins. Assim, se este algoritmo fosse utilizado na realidade, cinquenta e quatro pessoas que, em tese, deveriam receber crédito em conformidade com os critérios adotados pela instituição financeira, não receberiam.

Paralelamente, dos oitenta e seis clientes rotulados como maus pagadores na base de dados, o mesmo algoritmo acertou em cinquenta e dois casos, errando quanto aos restantes. Ao final, conclui-se que o algoritmo de Regressão Logística teve uma precisão de 70%.

Para o método de aprendizado de máquina Rede Neural Artificial, foi alcançada uma precisão de 75%, pouco superior que ao obtido pela Regressão Logística. Em relação à matriz de confusão, pôde-se perceber que o método rotulou corretamente cento e setenta e um bons pagadores, errando em quarenta e três casos. Quanto aos pagadores ruins, foram classificados corretamente cinquenta e seis clientes, com trinta erros.

4 Discussão

Os resultados obtidos, como se nota, não apontaram uma grande disparidade entre os métodos adotados. Dentre as possíveis explicações para isso, talvez a mais importante seja a base de dados, que é muito pequena (só possui dados de mil clientes) se comparada às utilizadas por instituições com milhares ou milhões de clientes.

Ainda assim, para os fins deste estudo, os resultados trazem alguns pontos interessantes.

Em primeiro lugar, é preciso considerar que revelar o banco de dados utilizado para o treinamento e o modelo decisório empregado caracterizaria uma clara afronta aos segredos comercial e industrial, já que seria fácil, a partir daí, criar um sistema muito parecido, ou mesmo idêntico, ao original. Sendo assim, um direito individual à explicação não poderia recair sobre tais informações.

Seria possível estabelecer, por outro lado, que o direito à explicação restaria atendido se fossem revelados os atributos utilizados e os pesos atribuídos a cada um deles. Neste caso, porém, existe um problema fundamental: não se pode considerar que os pesos atribuídos por um modelo de aprendizado de máquina seriam iguais, ou mesmo equivalentes, àqueles estabelecidos por um humano.

No exemplo aqui tratado, as variáveis que mais contribuem para o resultado final são: *status* da conta corrente, duração do empréstimo, histórico de crédito, saldo em conta poupança e o motivo do pedido de empréstimo. Embora não haja dúvidas de que estas variáveis são relevantes em um processo decisório do tipo, existem outros atributos considerados na base de dados que poderiam ser considerados tão ou mais relevantes que estes cinco, como o fato de o cliente estar empregado ou desempregado ou o fato de possuir imóveis em seu nome ou não.

Além disso, a revelação dos pesos exatos de cada uma das variáveis poderia também

eventualmente levar a uma violação dos segredos comercial e industrial, já que facilitaria a reprodução dos resultados.

Vale ressaltar, neste ponto, que, embora na Regressão Logística seja relativamente fácil identificar com exatidão os pesos atribuídos a cada um dos atributos, o mesmo não se pode dizer de modelos mais complexos como as Redes Neurais Artificiais, sendo que, nestes casos, a tarefa dependeria de ferramentas como LIME (RIBEIRO et al., 2016) ou SHAP (LUNDBERG; LEE, 2017).

Considerando as dificuldades aqui expostas, talvez seja importante determinar o grau de arbitrariedade permitido ao agente que realiza o tratamento de dados. No caso da análise de risco de crédito, uma instituição financeira, agindo como ente privado, parece ter um elevado grau de arbitrariedade em sua decisão, desde que, é claro, não adote práticas vedadas em lei, como as que gerem discriminação ou que importem na supressão de direitos individuais. Sendo assim, do ponto de vista individual, a revelação de pesos de atributos não parece ser necessária para fins de se determinar o alcance do direito à explicação conferido pela lei ao titular de dados pessoais, ao menos em tarefas para as quais há grande liberdade decisória.

Porém, é preciso considerar que existem tarefas em relação às quais a fundamentação do processo decisório é tão importante quanto a decisão em si. Para este tipo de tarefa, é importante pensar em um direito à explicação mais abrangente, que compreenda, por exemplo, algumas informações quanto ao modelo utilizado para a decisão ou quanto aos pesos atribuídos às variáveis consideradas.

5 Considerações Finais

Este estudo consiste em um levantamento preliminar acerca da abrangência do direito à explicação. Tomado sob uma perspectiva individual, um levantamento inicial sugere que tal direito deve ser mais ou menos abrangente a depender do grau de arbitrariedade permitido para que um determinado agente tome decisões automatizadas com base no tratamento de dados pessoais.

Vale ressaltar, porém, que este estudo não se aprofundou em questões relacionadas às diferentes técnicas hoje utilizadas para interpretar “sistemas de caixa preta”, sendo certo que, além de SHAP e LIME, já mencionados, existem ainda os modelos baseados em contrafactuais (MOLNAR, 2020; WACHTER et al., 2018).

Além disso, o problema da revisão de decisões automatizadas não foi tratado. A possibilidade aberta pela LGPD de a revisão de decisões automatizadas ser, ela mesma, também automatizada levanta a seguinte questão: se existe um sistema melhor ou mais justo que o original para tomar determinada decisão, por que este sistema simplesmente não substitui o original, eliminando a necessidade de revisão? Isso porque tanto o sistema utilizado para a revisão quanto o sistema que tomou a decisão em primeiro lugar serão determinísticos, de modo que suas decisões, para um cliente cujos atributos não variaram entre uma e outra, serão exatamente iguais.

Assim, o assunto merece maior aprofundamento científico.

Como futuras atividades de pesquisa, pretende-se tratar, de forma mais detida, destes pontos, a fim de que se possa contribuir para o estabelecimento de um regime de explicação, revisão e auditabilidade de sistemas automatizados de tomada de decisões.

Referências

BRASIL. Lei n° 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 15 out. 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>. Acesso em: 13 jun. 2020.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. Smote: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321–357, 2002.

HOFMANN, H. Statlog (german credit data) data set. **UCI Repository of Machine Learning Databases**, 1994.

JOHANSSON, U.; SÖNSTRÖD, C.; NORINDER, U.; BOSTRÖM, H. Trade-off between accuracy and interpretability for predictive in silico modeling. **Future medicinal chemistry**, Future Science, v. 3, n. 6, p. 647–663, 2011.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: **Proceedings of the 31st International Conference on Neural Information Processing Systems**. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 4768–4777. ISBN 9781510860964.

MOLNAR, C. **Interpretable Machine Learning: A Guide for Making Black Box Models Explainable**. [S.l.]: https://christophm.github.io/interpretable-ml-book/, 2020. <https://christophm.github.io/interpretable-ml-book/>.

MONTEIRO, R. L. Existe um direito à explicação na lei geral de proteção de dados do brasil? **Artigo estratégico 39**, Instituto Igarapé, Dez. 2018.

MORI, T.; UCHIHIRA, N. Balancing the trade-off between accuracy and interpretability in software defect prediction. **Empirical Software Engineering**, Springer, v. 24, n. 2, p. 779–825, 2019.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: Machine learning in python. **The Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. "why should i trust you?": Explaining the predictions of any classifier. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 1135–1144. ISBN 9781450342322. Disponível em: <https://doi.org/10.1145/2939672.2939778>.

WACHTER, S.; MITTELSTADT, B.; RUSSELL, C. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. **Harvard Journal of Law and Technology**, v. 31, p. 841–887, 04 2018.