

OPEN SOURCE INTELLIGENCE: CLASSIFICAÇÃO E MITIGAÇÃO DE RISCOS E FRAUDES NO ÂMBITO DE INSTITUIÇÕES FINANCEIRAS

Resumo — O alto volume de dados que trafegam nas comunicações abertas é uma realidade crescente. Por isso é necessário o aprimoramento da captação, análise e blindagem das informações, por meio da doutrina de Open Source Intelligence – OSINT. O objetivo do estudo é descobrir as vantagens do uso da OSINT pelas instituições financeiras e como mecanismos do atual Open Finance e LGPD podem contribuir no auxílio para melhor classificação de pontuação dos clientes, visando mitigar riscos, fraudes e maior eficácia na tomada de decisão. Como resultados preliminares do recorte utilizado, foi proposto um macromodelo para auxiliar na validação por parte das instituições financeiras dos seus respectivos modelos. O caso selecionado tem volume médio de € 100.000,00 (cem mil euros) por dia. Logo, justifica-se que a referida proposta permitirá independência às instituições, as quais terão domínio das informações utilizadas na classificação universal, além de resultados de outras informações pesquisadas em bases seguras.

Palavras Chave – Open Banking; OSINT; Data Mart; Credit score.

Abstract — The high volume of data that travels in open communications is a growing reality. Therefore, it is necessary to improve the capture, analysis and shielding of information through Open-Source Intelligence Technology–OSINT. The objective of this study is to uncover the benefits of using OSINT by financial institutions and how the mechanisms of the current Open Finance and LGPD can contribute to best classify customer scores, aiming to mitigate risks and fraud, and increase efficiency in decision-making. As preliminary results of the adopted outline, a macro model was proposed to help financial institutions validate their respective models. The selected case has an average daily volume of €100,000.00 (one hundred thousand euros). Thus, it is justified that the referred proposal will allow independence to the institutions, which will have control of the information used in the universal classification, in addition to the results of other information searched in secure bases.

Keywords – Open Banking; OSINT; LGPD; Credit score.

I. INTRODUÇÃO (PROBLEMÁTICA)

As instituições financeiras classificam seus clientes de acordo com regramento e política interna, tomando como base as informações trafegadas internamente em seus sistemas corporativos. Nessa conjuntura, a classificação pode divergir da realidade, em que, ao visar maior competitividade e

sobrevivência na era da informação, vislumbra-se a soma da tecnologia *Open Banking* e da doutrina *Open Source Intelligence* (OSINT) na geração de segunda camada para penalizar maus pagadores.

A implantação do *Open Banking* no Brasil, em novembro de 2020, representou uma verdadeira evolução no sistema financeiro nacional e trouxe benefícios para uma série de áreas, dando mais autonomia aos consumidores, à contabilidade, passando pelo surgimento de novos modelos de negócio que aumentam o dinamismo da economia brasileira e global [1]. Com o replanejamento da Lei Geral de Privacidade de Dados (LGPD) para agosto de 2021, as instituições financeiras têm a oportunidade de se adequar à norma.

Nesse contexto, justifica-se a análise da aplicabilidade no compartilhamento de dados sensíveis para o *Open Banking* e a proposição de requisitos de segurança envolvidos, com enfoque na LGPD, e advento da doutrina OSINT. A Inteligência de Fonte Aberta (*Open Source Intelligence* - OSINT) constitui na coleta, no processamento e na correlação de informações públicas de diversas fontes de dados abertos, como mídia de massa, redes sociais, fóruns, blogs, jornais, dados públicos do governo, entre outros [2]. Utilizando alguns dados de entrada, com emprego de técnicas avançadas de coleta e análise, a OSINT justifica as pesquisas de forma contínua para aprofundamento no assunto [3], prezando pela segurança e pela confiabilidade, evitando violações de segurança que eventualmente podem ser penalizadas.

De maneira geral, os autores convergem nos benefícios do uso da OSINT quanto à grande quantidade de informações, à alta capacidade de computação, *big data* e aprendizagem de máquina, como fonte de dados complementares aos dados já trafegados pelas instituições financeiras que aderiram ao *Open Banking*. Além disso, o estudo evidencia que o tratamento da informação deve ser realizado desde sua origem até as consultas nas *Application Programming Interface* (APIs) dos órgãos governamentais e das instituições financeiras [4]. Dessa forma, vislumbra-se menor risco a possíveis fraudes e má classificação dos seus clientes. Dessarte, este trabalho aborda um recorte que gira em torno de pelo menos 500 transações/dia, com valores

somados em torno de € 100.000,00 (cem mil euros).

Ante o exposto, o presente trabalho tem por objetivo o enriquecimento das informações coletadas no *Open Banking*, agregando estas a outras informações com uso de doutrina OSINT, abarcando uso da LGPD para estruturação de score universal seguro e confiável, podendo ser utilizado por instituições financeiras.

Para tanto, primeiramente serão apresentados os conceitos de *Open Banking*, OSINT e LGPD. Em seguida, a problemática será examinada e a solução apresentada, conforme modelo de penalização proposto. Por fim, serão abordados os prováveis desdobramentos do estudo.

II. OPEN BANKING, OSINT E LGPD

O score de crédito é uma pontuação que reflete o comportamento do consumidor. Por meio do score dos clientes, as empresas avaliam o quão arriscado é conceder o crédito a eles com base nos seus respectivos hábitos financeiros. Quanto mais alta a pontuação de crédito, melhor e mais fácil será o acesso a empréstimos, financiamentos, cartões e a contratação de qualquer tipo de crédito. Por outro prisma, quanto menor a pontuação, maior será a dificuldade na aprovação de crédito por parte das instituições financeiras.

O *Open Banking*, ou Sistema Financeiro Aberto (SFA), é uma iniciativa do Banco Central do Brasil que tem como principais objetivos trazer inovação ao sistema financeiro, promover a concorrência e melhorar a oferta de produtos e serviços financeiros para o consumidor. O *Open Banking* no mundo ainda está no início de suas atividades, mas modelos já em funcionamento, como no Reino Unido há 4 anos, por exemplo, servem de grande inspiração para a consolidação deste sistema no Brasil [5].

O consumidor é quem decidirá quando e com quem deseja compartilhar seus dados. Isso tornará o ecossistema financeiro mais competitivo e saudável para o consumidor final e suas necessidades.

A transferência de dados facilitada pelo *Open Banking* (apenas bancos) e, futuramente, pelo *Open Finance* (além dos bancos tradicionais, *fintechs*, corretoras e companhias de seguro, entre outros), dá aos consumidores acesso a um quadro financeiro mais completo, visando melhorar seus resultados, como empréstimos, investimentos e transações [1].

Esse sistema começou a ser implementado no Brasil no começo de 2021, mas o *Open Banking* no mundo já tem iniciativas prévias, como Austrália, Hong Kong, EUA e Singapura, sendo que grande parte da regulação brasileira foi inspirada no exemplo do Reino Unido, considerado pioneiro. Inclusive, neste ano, o Banco Central definiu que agora essa iniciativa se chama *Open Finance*, substituindo o *Open Banking*. A mudança, que é apenas nominal, facilita o entendimento do público e reforça a ideia de sistema financeiro aberto, justamente porque o *Open Finance* inclui produtos bancários tradicionais e serviços financeiros, como câmbio, seguros e previdência [5].

A OSINT, como qualquer outro tipo de inteligência, deve preconizar uma metodologia bem estabelecida [6]. A inteligência obtida de fontes governamentais abertas é de suma importância para o complemento das informações e a tomada de decisão. Essa técnica de coleta, análise e tratamento da informação é disciplinada como OSINT. O pesquisador pode utilizar do melhor método para a extração e o trabalho da informação para atingir o objetivo esperado, sendo uma boa prática verificar as fontes, higienizando-a e valorando-a quanto à finalidade esperada [2].

Um dos exemplos de desafios é a questão da automação do processo de coleta, pois quanto maior a quantidade de informações coletadas, maior a probabilidade de criar inferências e relacionamentos. No entanto, a quantidade de dados disponíveis hoje é muito vasta, sendo, assim, impossível ser coletada de forma manual.

Dos benefícios amplamente explorados pelo estado da arte, destaca-se sua utilização por governos e serviços de inteligência para condução de investigações e combate ao crime cibernético [7], além de ser utilizada para marketing, campanha política ou gestão de desastres [8]. Das desvantagens, algumas associações mal-intencionadas podem tirar proveito dessas informações divulgadas no formato on-line para atrair, de alguma forma, seu alvo e capturá-lo e para a execução de avaliações cognitivas automáticas da veracidade dos dados abertos, visando à divulgação de *fake news* e *deep fakes* [3]. Segundo os autores [4], há maior confiabilidade dos dados e segurança na tomada de decisão quando estes são armazenados de forma distribuída e em maior granularidade.

LGPD é a primeira legislação brasileira que regulamenta amplamente a privacidade de dados. Tem por objetivo regulamentar o tratamento de dados pessoais garantindo os direitos fundamentais de proteção da liberdade e da privacidade das pessoas. Além disso, fornece aos titulares dos dados mais transparência e controle sobre a coleta e o uso de seus dados. Também harmoniza as leis de privacidade de dados em todos os estados do Brasil. A LGPD e o Regulamento Geral de Proteção de Dados (GDPR) da União Europeia são muito próximos.

Um conceito importante a considerar é a definição de dados pessoais, que são todas as informações (físicas ou digitais) relacionadas a pessoas físicas identificadas ou identificáveis e dados pessoais. Outro conceito são dados pessoais sensíveis relacionados à origem racial ou étnica, crença religiosa, opinião política, filiação sindical, organização religiosa, filosófica ou política, dados de saúde ou vida sexual, dados genéticos ou biométricos e pessoa física. O titular dos dados é a pessoa singular a quem remete os dados pessoais objeto de tratamento [17].

A estrutura da LGPD foi baseada em um modelo conceitual, identificando seus principais componentes e suas inter-relações sem perder a consistência de sua essência. Além disso, deve ser aberto e flexível, permitindo a adição de novos conteúdos ou componentes sem prejudicar sua integridade e consistência. Por

fim, ele deve estar alinhado com os principais padrões, diretrizes e regulamentos globais de TI.

III. PROBLEMA E PROPOSTA DE SOLUÇÃO

A avaliação de crédito é um dos processos mais cruciais nas decisões de gestão de crédito dos bancos. Esse processo inclui coletar, analisar e classificar diferentes elementos e variáveis de crédito para criar um indicador - o score de crédito - que permita aos bancos avaliarem decisões de crédito de modo a prevenirem a contratação de empréstimos por maus pagadores.

Esse score de crédito, por sua vez, é uma estimativa baseada em modelo da probabilidade de que uma pessoa que tomou um empréstimo irá mostrar algum comportamento indesejável no futuro [10]. Um processo base para o desenvolvimento de modelo de score de crédito é apresentado na Figura 1.

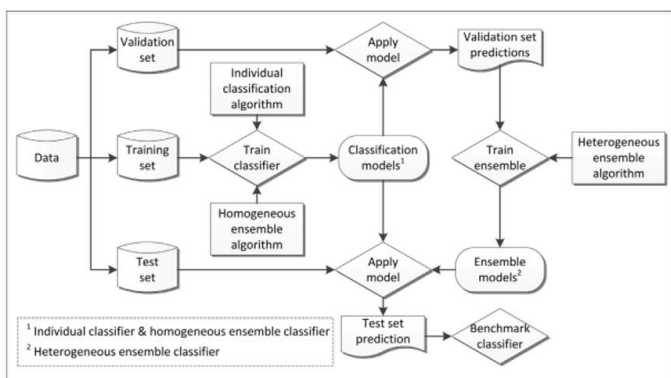


Figura 1. Processo exemplo para desenvolvimento de um modelo de score de crédito [10].

Para a construção dos modelos de score de crédito, são várias as técnicas possíveis, a exemplo de árvores de decisão [11], k-vizinhos mais próximos [12], redes neurais [13] e máquinas de vetores de suporte (SVMs) [14].

Apesar de muita pesquisa, observa-se que a literatura de pontuação de crédito não reflete vários avanços recentes na aprendizagem, apesar do trabalho desenvolvido por [15] comprovar que o desempenho de modelos de pontuação baseados em aprendizado de máquina melhorou substancialmente desde a adoção dos algoritmos baseados em comitês, a exemplo dos trabalhos propostos por [10] e [16].

As instituições financeiras utilizam, em sua maioria, seus próprios algoritmos, políticas e regras na montagem do score dos clientes, utilizando somente as informações contidas em suas bases. Essa realidade pode gerar um score pouco realista, no qual a instituição financeira perde negócio com clientes em potencial, ou pior, faz negócio com clientes bem pontuados, mas que não condiz com a realidade. Essa proposta tende a fazer um score universal e compartilhado entre as instituições financeiras.

O processo de montagem do score de crédito por parte dos bancos brasileiros se baseia na individualização de pessoas, físicas ou jurídicas, com a utilização do Cadastro de Pessoas

Físicas (CPF) e do Cadastro Nacional de Pessoas Jurídicas (CNPJ). Associado a esse atributo principal, observam-se, ainda, o uso dos seguintes componentes na montagem do modelo:

- A. Verificação de renda e emprego: verifica a renda do cliente e sua fonte de renda (por exemplo, salário, renda de aluguel, pensão, etc.), a fim de determinar a capacidade de pagamento do empréstimo. Além disso, verifica se o cliente está empregado ou não e há quanto tempo está no emprego atual.
- B. Avaliação do histórico de crédito: avalia o histórico de crédito do cliente, verificando se ele ou ela já tem outros empréstimos ou financiamentos em andamento, bem como se tem um bom histórico de pagamentos. Também verifica se o cliente já teve problemas de inadimplência, como atrasos no pagamento de dívidas.
- C. Análise do patrimônio: verifica se o cliente tem patrimônio que possa ser usado como garantia do empréstimo, como imóveis ou veículos. Isso ajuda a reduzir o risco do empréstimo, já que a instituição financeira pode tomar posse desses bens em caso de inadimplência.
- D. Cálculo da capacidade de pagamento: com base nas informações coletadas, é calculada a capacidade de pagamento do cliente, ou seja, quanto ele ou ela pode pagar por mês sem comprometer sua renda. Isso ajuda a determinar qual é o valor máximo do empréstimo que pode ser concedido.
- E. Com base em todas essas informações, é avaliado o risco de conceder o empréstimo ao cliente. Se o risco for considerado muito alto, o banco pode optar por não conceder o empréstimo ou oferecer uma taxa de juros mais alta para compensar o risco.
- F. Por fim, a instituição financeira toma a decisão se deve conceder o empréstimo ou não, com base em todos os fatores acima. Se o empréstimo for concedido, estipula-se os termos e condições, como prazo de pagamento e taxa de juros, e o cliente pode aceitar ou não.

Como apresentado, muito do que é utilizado na montagem do score é derivado de atributos internos aos bancos, o que pode levar a um enviesamento do modelo. Ao mesmo tempo, uma série de outros atributos relacionados a pessoas físicas e jurídicas, oriundos de dados abertos governamentais, disponíveis para acesso público e que poderiam melhorar a acurácia do modelo, são desconsiderados.

Como exemplos de atributos relevantes para um modelo de classificação de crédito, poderíamos citar as bases de dados de dívida ativa da União, estados e municípios – que armazenam informações de devedores na esfera pública, ou de empresas inidôneas, suspensas ou punidas – que traz à luz dados de empresas com mau relacionamento com o governo brasileiro,

além de informação provida por outros bancos por meio do *Open Banking*.

É importante ressaltar que muito do que se observa na dificuldade dos bancos em promover mudanças no seu modelo vem do fato de que, em grande parte das instituições, a tecnologia utilizada é proprietária, ou seja, é oriunda da expertise do fornecedor do serviço, associado ao risco real de que uma mudança brusca do modelo possa levar a resultados inconsistentes ou indesejados, o que, em casos extremos, poderia até mesmo levar a instituição a falência.

Por esse motivo e de forma a viabilizar a aplicação prática desta proposta, optou-se por não alterar o modelo base de cálculo do score, mas pela criação de uma segunda camada, que seria responsável por aplicar uma penalidade sobre a nota já calculada. Nessa linha, a intensidade da penalidade imposta por cada atributo seria estabelecida por pesos calculados com base em um conjunto anotado de registro de concessões de empréstimos, de acordo com o apresentado na Figura 2.

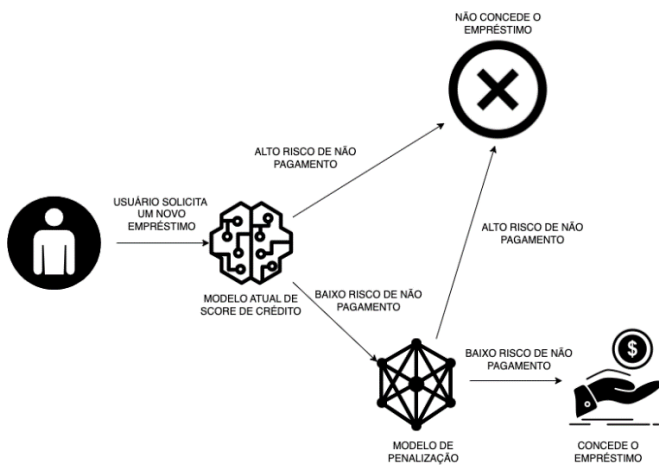


Figura 2. Proposta do macromodelo de penalização associado do modelo de score de crédito bancário.

É importante ressaltar que a implementação da etapa de penalização impactaria positivamente na negação de empréstimos a maus pagadores, uma vez que os atributos associados às bases de origem traduzem apenas características negativas relacionadas a pessoas físicas ou jurídicas sob análise, levando a uma penalização que sempre irá descontar pontos do proponente do empréstimo. Por outro lado, é provável que, dentre aqueles afetados com a negativa de empréstimo, o sistema leve a inclusão, de forma incorreta, de prováveis bons pagadores.

Por fim, o que se espera, de forma geral, é que o balanço do prejuízo evitado com concessões indevidas de empréstimos a maus pagadores em função do lucro que deixou de ser aferido com a negativa de empréstimo a possíveis bons pagadores seja positivo.

IV. MODELO DE PENALIZAÇÃO

Internamente, o modelo de penalização se baseia na definição de pesos para cada um dos atributos coletados nas bases governamentais dos tribunais de contas (que armazenam os dados de devedores públicos) e nos registros de empresas inidôneas, suspensas e punidas. Dessa forma, para cada novo usuário, serão gerados 30 novos atributos numéricos associados:

- 28 atributos referentes a registros de débitos em cada uma das 27 unidades da Federação e em relação à União.
- 1 atributo referente à suspensão na CGU.
- 1 atributo referente à inidoneidade na CGU.
- 1 atributo referente à penalidade na CGU.

Para cada um dos atributos, será calculada a quantidade de registros para um dado usuário, vez que um mesmo usuário pode ter múltiplos registros de débitos junto a cada ente. A cada um desses atributos o modelo atribui um peso inicial aleatório. O ajuste dos pesos se dá a partir da etapa de treinamento do modelo, à medida que registros anotados de empréstimo são alimentados e o resultado final da classificação é obtido; a cada ocorrência de erro, os pesos são recalculados, até que o modelo venha a convergir (Figura 3). Ao final do processo, espera-se que os atributos mais relevantes mantenham pesos mais altos, ao tempo que os menos significantes possam até mesmo ser descartados – peso 0.

Para o recálculo dos pesos, é proposta a utilização de uma estratégia similar a utilizada em redes perceptron de camada única, com a inclusão de um elemento de viés (bias) e uma função de ativação, conforme apresentado na Figura 3.

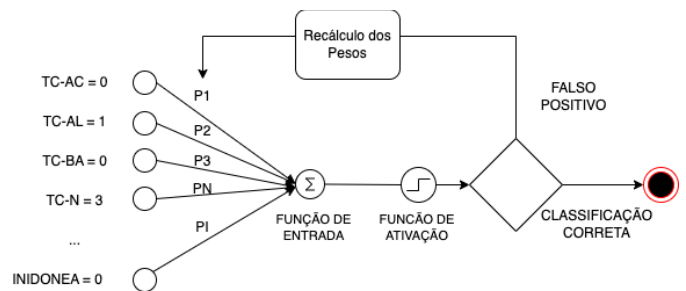


Figura 3. Detalhamento do modelo proposto.

V. CONCLUSÃO E TRABALHOS FUTUROS

Com uso da nova proposta, modelo de penalização, baseada em agregar informações utilizando *Open Banking* e OSINT, todo trabalho de levantamento e mapeamento do processo negocial a ser automatizado deve gerar uma base de conhecimento sólida e uma validação por parte das instituições financeiras dos seus respectivos modelos de score de clientes, em que todos os questionamentos e apontamentos são resolvidos.

É importante ressaltar que o presente estudo ainda está em desenvolvimento, de modo que a implementação e a validação

do modelo teórico apresentado, bem como das ferramentas descritas, ainda não foi concluída. A completude da informação traz diferencial competitivo com melhor classificação dos clientes, mitigação dos riscos de imagem e possíveis fraudes, como lavagem de dinheiro por parte de clientes mal-intencionados, analisando dados abertos de fontes governamentais seguras, *Open Banking* e registrando tais informações em um Data Mart.

Nessa temática, observa-se que a proposta pode avançar com outras possibilidades, as quais serão testadas a posteriores com a utilização do OSINT, fazendo buscas em fontes de dados abertos de mídia social, além do consumo de outras APIs abertas, como cartórios (pesquisar imóveis) e Detrans (pesquisar veículos), enriquecendo e agregando valor a ferramentas adquiridas e dando maior segurança e completude da informação.

Como se trata de uma proposta inicial, conclui-se que o presente estudo atingiu os resultados relacionados à geração de um macromodelo, visando à melhor classificação dos clientes, podendo servir de base universal para validação dos modelos das instituições financeiras. No entanto, é necessário aplicar o modelo proposto e, além disso, aplicar em uma amostra maior de dados, a fim de validar e checar os resultados frente ao problema, recomendando-se aprimorar as especificações de desenvolvimento da solução, requisitos de desempenho, segurança, interface de usuários, operacionais, pós adequação das instituições financeiras, junto ao órgão regulador, BACEN, para verificação dos resultados e ganhos diretos e indiretos.

REFERÊNCIAS

- [1] NACIONAL, I. **RESOLUÇÃO CONJUNTA Nº 1, DE 4 DE MAIO DE 2020** - DOU - Imprensa Nacional. 2020. Disponível em: <http://www.in.gov.br/web/dou>.
- [2] TANABE, R. *et al.* **OSINT Methods in the Intelligence Cycle**. Professional Post-Graduate Program in Electrical Engineering. 2022.
- [3] JIN, Zhiwei *et al.* News verification by exploiting conflicting social viewpoints in microblogs. In: **Proceedings of the AAAI conference on artificial intelligence**. 2016.
- [4] ALVES, D. *et al.* **Sustainable Cities and Society**. Post-Graduate Program in Electrical Engineering. 2021. IoT-based security service for the documentary chain of custody, *Sustainable Cities and Society*, Volume 71, 2021,102940.
- [5] BRODSKY, L.; OAKES, L. **Capitalizing on the potential benefits of open banking** | Mc-Kinsey. 2017. Disponível em: <https://www.mckinsey.com/industries/financial-services/our-insights/data-sharing-and-open-banking>.
- [6] TABATABAEI, Fahimeh; WELLS, Douglas. OSINT in the Context of Cyber-Security. **Open source intelligence investigation**, p. 213-231, 2016.
- [7] NOUH, M; NURSE, J. R.; WEBB, H; M. GOLDSMITH, M. **Cybercrime investigators are users too! Understanding the socio-technical challenges faced by law enforcement**, in Proc. 2019 Workshop Usable Security, Feb. 2019.
- [8] POWELL, A.; C. HAYNES, C. "Social media data in digital forensics investigations," in **Digital Forensic Education: An Experiential Learning Approach**, X. Zhang and K.-K. R. Choo, Eds. Cham, Switzerland: Springer, 2020, pp. 281–303.
- [9] RAMOS, J.; ALTURAS, B; MORO, S. "Business intelligence in a public institution — Evaluation of a financial data mart," **2017 12th Iberian Conference on Information Systems and Technologies (CISTI)**, Lisbon, Portugal, 2017, pp. 1-6, doi: 10.23919/CISTI.2017.7975672.
- [10] LESSMANN, Stefan *et al.* **Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research**. *European Journal of Operational Research*, v. 247, n. 1, p. 124-136, 2015.
- [11] COFFMAN, J. Y. **The proper role of tree analysis in forecasting the risk behavior of borrowers**. *Management Decision Systems*, Atlanta, MDS Reports, v. 3, n. 4, p. 7, 1986.
- [12] HENLEY, W. E. *et al.* **Construction of a k-nearest-neighbour credit-scoring system**. *IMA Journal of Management Mathematics*, v. 8, n. 4, p. 305-321, 1997.
- [13] YOBAS, Mumine B.; CROOK, Jonathan N.; ROSS, Peter. **Credit scoring using neural and evolutionary techniques**. *IMA Journal of Management Mathematics*, v. 11, n. 2, p. 111-125, 2000.
- [14] BAESENS, Bart *et al.* **Benchmarking state-of-the-art classification algorithms for credit scoring**. *Journal of the operational research society*, v. 54, p. 627-635, 2003.
- [15] DUMITRESCU, Elena *et al.* **Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects**. *European Journal of Operational Research*, v. 297, n. 3, p. 1178-1192, 2022.
- [16] FINLAY, Steven. **Multiple classifier architectures and their application to credit risk assessment**. *European Journal of Operational Research*, v. 210, n. 2, p. 368-378, 2011.
- [17] BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados (LGPD) Brasília: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/13709.htm. Acesso em: 4 abr. 2023.