

ENAP

**CIÊNCIA DE DADOS EM
POLÍTICAS PÚBLICAS:
UMA EXPERIÊNCIA DE
FORMAÇÃO**

Especialização em
Ciência de Dados
aplicada a
Políticas Públicas

Ciência de Dados em Políticas Públicas: uma experiência de formação

Organizadores:

Jackson De Toni e Rachel Dorneles

Colaboradores:

Natália Massaco Koga, Alex Lopes Pereira e

Hélio Bomfim de Macêdo Filho

Autores:

Alexandre Martins Nogueira, Ciro Jônatas de Souza Oliveira, Elias Fernandes de Oliveira, Erica Ramalho de Oliveira, Glauber Volkmer, Gustavo Queiroz Chaves, Isvaldo Silveira Lima Filho, José Leonardo Ribeiro Nascimento, Leonardo Lins da Câmara Marinho Barbalho, Marcos Cândido de Paula Rezende, Matheus Scatolino de Rezende, Mônica Rondina, Rafael Leal Garcia, Rafaela de Sá Gonçalves, Rodrigo Santos de Paula, Romualdo Anselmo dos Santos, Ronan Luiz da Silva, Sérgio Luiz Gomes de Carvalho e Thiago Morais Furuchima



Escola Nacional de Administração Pública – Enap

Presidente

Diogo Costa

Diretor de Educação Executiva

Rodrigo Torres

Coordenadora-Geral de Especialização e MBA

Bruna Danielly da Paz Tenório

Coordenadora de Especialização e MBA

Juliana Miranda Leda

Coordenadora do curso

Rachel Pereira Dorneles

Coordenadora do Projeto

Renata Regina Scarpim

Equipe da Coordenação

Cecília Baran, Eliane Luz, Teresa Soares, Leonardo Batista, Gustavo Sousa, Mário Souza e Genival Macedo | Lucas Oliveira (Estagiário)

Organizadores

Jackson De Toni, Rachel Dorneles

Colaboradores

Natália Massaco Koga, Alex Lopes Pereira, Hélio Bomfim de Macêdo Filho

Revisão

Angélica Magalhães Turquetti

Projeto gráfico e diagramação

Ana Carla Gualberto Cardoso

A Escola Nacional de Administração Pública (Enap) é uma escola de governo vinculada ao Ministério da Economia (ME).

Tem como principal atribuição a formação e o desenvolvimento permanente dos servidores públicos. Atua na oferta de cursos de doutorado e mestrados profissionais, especialização *lato sensu*, cursos de aperfeiçoamento para carreiras do setor público, educação executiva e educação continuada.

A instituição também desenvolve pesquisa aplicada e ações de inovação voltadas à melhoria do serviço público.

Ciência de Dados em Políticas Públicas: uma experiência de formação

ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS
APLICADA A POLÍTICAS PÚBLICAS

Ficha catalográfica elaborada pela equipe da Biblioteca Graciliano Ramos da Enap

C5741 Ciência de dados em políticas públicas: uma experiência de Formação / organizado por Jackson de Toni e Rachel Dorneles; colaboração de Natália Koga, Hélio Macedo e Alex Pereira. -- Brasília: Enap, Especialização em Ciência de Dados aplicada a Políticas Públicas, 2022.

269 p. : il.

Inclui bibliografia.

ISSN: 978-65-87791-30-2

1. . Ciência de Dados. 2. Políticas Públicas. 3. Inovação. 4. Tecnologia da Informação. I. Título. II. Toni, Jackson de. III. Koga, Natália. IV. Macedo, Hélio. V. Pereira, Alex

CDD 320.6

Bibliotecária: Tatiane de Oliveira Dias – CRB1/2230



Enap, 2022

Este trabalho está sob a Licença *Creative Commons* – Atribuição: Não Comercial – Compartilha Igual 4.0 Internacional

As informações e opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista da Escola Nacional de Administração Pública (Enap). É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

Escola Nacional de Administração Pública (Enap)

Diretoria de Educação Executiva

Coordenação-Geral de Especialização e MBA

SAIS – Área 2-A – 70610-900 — Brasília-DF, Brasil

Sumário

Agradecimentos	7
Apresentação.....	9
<i>Rodrigo Torres</i>	
Prefácio.....	11
<i>Natália Koga</i>	
Introdução.....	17
<i>Alex Lopes Pereira, Hélio Bomfim de Macêdo Filho</i>	
Análise descritiva das despesas de tecnologia da informação dos órgãos do Poder Executivo Federal	22
<i>Alexandre Martins Nogueira</i>	
Pesquisa qualitativa e descritiva sobre base normativa nacional de IA à luz de experiências internacionais voltadas ao uso dessa tecnologia pelo poder público.....	44
<i>Gustavo de Queiroz Chaves</i>	
<i>Machine learning</i> aplicada à análise dos fatores determinantes da taxa de ocupação e conclusão no ensino superior no Brasil	88
<i>Ivaldo Silveira Lima Filho</i>	
Índice de priorização de objetos de auditoria: um estudo de caso para municípios sergipanos.....	108
<i>José Leonardo Ribeiro Nascimento</i>	
Classificação automática de compras licitadas pelo sistema Compras.gov.br	127
<i>Matheus Scatolino de Rezende</i>	

Avaliação da atuação dos bancos públicos como agentes de política pública na cobertura de falhas de mercado.....	147
<i>Mônica Rondina</i>	
Análise da aplicação da sanção do embargo no combate ao desmatamento ilegal na Amazônia.....	164
<i>Rafael Leal Garcia</i>	
Perfil dos municípios produtores de cana-de-açúcar em relação a queimadas ou a incêndios, à área colhida e à presença de produtor de biocombustível incentivado pelo programa RenovaBio.....	187
<i>Rodrigo Santos de Paula</i>	
Análise de sentimentos e modelagem de tópicos de mensagens do <i>twitter</i> como instrumento para monitoramento contínuo de políticas públicas: um estudo de caso da política pública do Sistema Único de Saúde.....	216
<i>Thiago Morais Furuchima</i>	
Análise do Cadastro Único (CadÚnico): um recorte de antes e depois da pandemia da covid-19	235
<i>Ciro Jonatas de Souza Oliveira, Glauber Volkmer, José Leonardo Ribeiro Nascimento, Leonardo Lins Camara Marinho, Marcos Cândido de Paula Rezende, Monica Rondina, Romualdo Anselmo dos Santos E Sérgio Luiz Gomes de Carvalho</i>	
Panorama da situação de insegurança alimentar e nutricional dos municípios brasileiros após a pandemia da covid-19	251
<i>Elias de Oliveira, Erica Ramalho de Oliveira, Gustavo de Queiroz Chaves, Ivaldo Lima Filho, Rafaela de Sá Gonçalves e Ronan Silva</i>	
Sobre os autores, organizadores e colaboradores	265
Depoimentos de alunos.....	269

AGRADECIMENTOS

A Coordenação do curso de Ciência de Dados aplicada a Políticas Públicas agradece imensamente a todas professoras e a todos professores que fizeram parte desta realização. A atuação de cada uma delas e de cada um deles evidencia uma parceria muito enriquecedora para todos nós envolvidos neste desafio, que não são os rotineiros. Foram esforços de contextualização dos conhecimentos que vão alimentando novos saberes, em diálogos altamente pedagógicos. Seus nomes e os respectivos componentes curriculares ministrados são os seguintes.

Antônio Lassance de Albuquerque Júnior — Análise de Políticas Públicas e Processo Decisório.

Leandro Couto — Desenho e Avaliação Ex-Ante de Políticas Públicas.

Natália Koga e Bruno Gontyjo — Políticas Públicas Baseadas em Evidências.

Armando Amorim Simões — Monitoramento e Avaliação Ex-Post de Políticas Públicas.

Alex Lopes Pereira — Introdução a Programação em Python e Introdução a Ciência de Dados.

Márcio Victorino — Introdução a Banco de Dados e Linguagem SQL.

Eduardo Monteiro de Castro Gomes — Estatística Descritiva.

James Matos Sampaio — Estatística Inferencial.

Sara Vieira — Design Sprint e Lean Inception.

Erick Muzart e Maxwell Sarmiento — Machine Learning.

Luiz Pacheco Motta — Sistemas de Informação Geográfica.

Nara Kohlsdorf — Diálogos Metodológicos.

Jackson De Toni e Rodrigo Moraes Lima Delgado — Disciplina Integradora Aplicada.

Fernanda Serpa, Jackson De Toni, Hélio Macedo — Laboratório de Casos de Políticas Públicas de Saúde.

Jeniffer Chaves, Jackson De Toni, Hélio Macedo, Natália Koga — Laboratório de Casos de Políticas Públicas Sociais.

Rodolfo Benevenuto, Maxwell Sarmiento, Jackson De Toni, Hélio Macedo — Laboratório de Casos de Políticas Públicas de Infraestrutura.

Stefano Mozart e Hélio Macedo — Monitoria em Ciência da Computação e Monitoria em Matemática e Probabilidade.

Peng Yaohao, Thiago Marzagão e Leonardo Monastério — Conhecimentos Conexos na aula inaugural

Maurício Lima Barreto — Conhecimentos Conexos na aula de encerramento.

Fernanda Rodrigues — Conhecimentos Conexos na dinâmica de avaliação intermediária do curso.

Lucas Hiroshi, Clayton Silva, Ronnie Dilli, Rodrigo Botelho, Thiago Gomes — Monitores no eixo temático de computação.

APRESENTAÇÃO

RODRIGO TORRES⁽¹⁾

Os últimos anos têm vivenciado uma revolução na forma como os dados são produzidos e como circulam entre pessoas, empresas e governos. As novas tecnologias existentes potencializaram a importância dos dados para a humanidade, e o mesmo ocorreu nos governos ao redor do mundo.

Essa nova realidade passou a exigir novas competências daqueles que trabalham nos governos hoje. Ao mesmo tempo, há um claro *gap* entre as competências necessárias e a capacidade existente nos seus recursos humanos.

Segundo o estudo do BID chamado “Transformación Digital y Empleo Público: El Futuro Del Trabajo Del Gobierno”, 51% dos 718 gerentes públicos latino-americanos pesquisados reconhecem ter um déficit severo ou muito severo em suas habilidades de análise de dados.

Em uma pesquisa realizada pela Enap em 2021, foi feita uma estimativa de que aproximadamente 20% da atual força de trabalho do serviço público federal brasileiro encontra-se em ocupações com elevado potencial de automatização nas próximas décadas.

Em 2020, foram lançadas pelo Governo Federal as 7 competências transversais de um setor público de alto desempenho, incluindo “resolução de problemas com base em dados”. Isso indica que todo e qualquer servidor deve desenvolver essa competência para ter um bom desempenho.

Diante desses diagnósticos, a Enap vem oferecendo um conjunto de cursos e ações de formação para contribuir fortemente com o desenvolvi-

mento de capacidades relacionadas ao processo de aquisição, manipulação, processamento e análise de dados, gerando novos conhecimentos e novas possibilidades para a solução de problemas complexos no Governo Federal.

Diante desses desafios, e tendo a inovação e a geração de valor público impressa em seu DNA, a Enap tem buscado trazer discussões e debates de ponta nessa temática.

Atualmente, além de cursos regulares de aperfeiçoamento de carreiras e de cursos de curta duração na Escola Virtual de Governo, são ofertados *bootcamps* em ciência de dados, são organizados eventos sobre o tema, além de ofertados serviços que auxiliam os órgãos de governo na sistematização de dados para a tomada de decisão.

Como consequência desse conjunto de ofertas, a Enap decidiu começar a desenvolver a Especialização em Ciência de Dados aplicada a Políticas Públicas, gerando oportunidade para que os servidores públicos pudessem desenvolver e aprimorar suas capacidades e competências profissionais, tornando-os capazes de aplicar os conhecimentos de Ciências de Dados na análise e avaliação de Políticas Públicas no âmbito do governo brasileiro em suas diferentes esferas.

O curso teve carga horária total de 375 horas, foi desenvolvido em parceria com a CGU e ofertou ao todo 40 vagas, incluídas vagas para pessoas autodeclaradas negras, indígenas e com deficiência.

Este livro é um grande resultado dessa Especialização, apresentando trabalhos de conclusão de curso com reflexões aplicadas aos problemas públicos que os alunos selecionaram.

É uma alegria e um orgulho poder entregar um livro contendo tão rico material sobre a nossa experiência e com os melhores resultados obtidos ao longo dessa especialização. Desejo que possam absorver um pouco do que foi este processo e que possamos, juntos, trabalhar para aumentar a capacidade daqueles que integram o serviço público em realizar melhores entregas à sociedade.

PREFÁCIO

NATÁLIA KOGA ⁽²⁾

Há mais de cinco décadas, a Ciência de Dados tem se desenvolvido em torno das ideias e das iniciativas voltadas ao *aprendizado por meio dos dados* (DONOHO, 2017). Configurando-se hoje como uma área do conhecimento multidisciplinar que vincula métodos matemáticos e estatísticos com técnicas computacionais para a modelagem de problemas em distintos campos, a Ciência de Dados traz consigo fortes apostas quanto às possibilidades de sua aplicação no contexto das organizações públicas, tema no qual se insere esta publicação.

Se, por um lado, os avanços tecnológicos nas últimas décadas têm permitido a intensificação expressiva da produção de dados e o desenvolvimento de métodos para analisá-los, por outro, identificamos a proliferação de problemas públicos complexos que demandam toda sorte de recursos capazes de ampliar a compreensão e de criar alternativas para lidar com tais problemas. É, na relação entre essas duas dinâmicas, que a Ciência de Dados desponta como potencial aliada no direcionamento do uso dessa massa espantosamente crescente de dados para a produção de mais e de melhores subsídios informacionais para

a ação estatal, ou, como tem se referido mais recentemente, para a produção de evidências que melhor informem as políticas públicas.

Vários benefícios podem ser apontados no uso da Ciência de Dados para subsidiar a tomada de decisão nas políticas públicas, tais como a expansão dos dados simultaneamente analisados a respeito de diferentes dimensões e áreas de política (FILGUEIRAS; LUI, 2022), a economicidade, a agilidade e a replicabilidade de seus modelos e, até mesmo, no compartilhamento do ambiente de análise com os *stakeholders* (grupos interessados na política), ampliando assim a participação e a comunicabilidade entre eles (FURTADO; LASSANCE, 2022). De fato, governos, em todo mundo, vêm utilizando recursos da Ciência de Dados para o aprimoramento dos serviços e das políticas públicas, como redução de riscos de fraudes ou evasão fiscal, avaliação de risco de pacientes em serviços de saúde, monitoramento de surtos de doenças infecciosas, simulações dos efeitos de danos no meio ambiente, previsão de pensões e aposentadorias, para citar alguns.

Diante de tantos atributos, a literatura especializada mais recente tem discutido as possíveis implicações epistemológicas das rápidas transformações geradas no campo para as práticas de pesquisa e para a construção do conhecimento (KITCHIN, 2014). É nesse sentido que duas reflexões merecem ser colocadas especialmente para os analistas de políticas públicas. A primeira delas diz respeito à cadeia de absorção de subsídios informacionais, muitas vezes longa e tortuosa, do processo de produção de políticas públicas. E a segunda, quanto ao papel da Ciência de Dados nessa cadeia.

O uso de dados e informações gerado pelos analistas não ocorre de forma imediata e direta pelos tomadores de decisão, ele depende de várias etapas e condições para que seja bem-sucedido. Para que os dados sejam produzidos e utilizados como evidências, é necessário que exista, nas organizações públicas, conhecimento prévio acumulado tanto sobre o tema público em questão, quanto sobre os métodos a serem empregados, assim como fontes de dados e infraestrutura de processamento e análise estabelecidas. Em outras palavras, a análise de dados não ocorre em ambiente vazio, em contrário, depende do acúmulo de conhecimento e das capacidades já instaladas.

Ademais, além da preparação, da modelagem e da apresentação dos dados, é importante ter em mente que há toda uma cadeia posterior de absorção das análises geradas que pode fazer com que elas sejam apropriadas, total ou parcialmente, ou mesmo não apropriadas como evidência para a tomada de decisão. Tal cadeia envolve um fluxo de várias etapas que vão desde o reconhecimento do valor dessa fonte de conhecimento, passando por iniciativas efetivas para adquiri-la, a sua interpretação à luz de outros conhecimentos pré-existentes (inclusive produzidos por outros tipos de fontes), e, por fim, seu uso efetivo que pode ter diferentes aplicações (OUITMET *et al.*, 2009). O uso pode se dar não apenas para o fim instrumental de resolução imediata de problemas ou modificação de um contexto, mas também para uma finalidade mais indireta de ampliação e acúmulo do conhecimento sobre uma questão pública ou mesmo um uso simbólico de justificação de uma decisão já tomada (AMARA *et al.*, 2014).

O campo de análise de políticas públicas, no mesmo desenrolar do tempo em que o campo de Ciência de Dados se desenvolve, tem debatido e demonstrado que a política pública não é feita apenas de racionalidade instrumental, e que o fator político e a necessidade de legitimação das decisões governamentais pela sociedade são constitutivos de um processo democrático de tomada de decisão (SIMON, 1956; LINDBLOM, 1959; FISCHER E GOTTWEIS, 2012; LEJANO, 2006; SPINK, 2019; KOGA *et al.*, 2022). Sendo assim, há sim que se reconhecer o papel da Ciência de Dados como meio cada vez mais presente e imprescindível para melhor informar a produção das políticas, mas igualmente relevante é considerar que não se trata da panaceia que sozinha resolverá todos os problemas públicos e que, como qualquer meio, também pode ser carregada de vieses e aplicada para diversas finalidades. Portanto, um olhar crítico e atento para que ela continue contribuindo, sem desconsiderar o conhecimento já produzido por outras fontes de evidência, de forma transparente e sempre aberta ao escrutínio público é fundamental para garantir o seu bom uso na produção das políticas.

O curso de *Especialização em Ciência de Dados aplicada a Políticas Públicas*, ofertado pela Enap, teve a preocupação, desde a sua concepção, de apresentar

as diferentes potencialidades e os desafios da aplicação da Ciência de dados. Portanto, além do desenvolvimento de um rol abrangente e atualizado de capacidades técnicas, também procurou contextualizar seus usos na dinâmica da produção das políticas públicas por meio de aulas expositivas, laboratórios aplicados e oficinas interdisciplinares que promoveram a reflexão crítica, aproveitando do conhecimento e da experiência dos profissionais envolvidos, sejam os próprios alunos e alunas, sejam os docentes. Esta publicação reflete essa abordagem multifacetada e integrada, assim como o esforço dos alunos, das alunas, dos orientadores e das orientadoras que buscaram traduzir seus aprendizados e suas interações em textos sintéticos.

Ainda que não apresente todos os trabalhos concluídos no curso por limitação de tempo e espaço, a publicação reúne trabalhos que garantem um bom retrato da diversidade de métodos apreendidos, como detalhado na Introdução, e da diversidade de finalidades para as quais a Ciência de Dados pode ser aplicada. Abarcam, portanto, usos mais comumente identificados com a análise de políticas públicas, como teste de modelos preditivos, mapeamento de demandas, avaliação de desempenho de entes públicos e de políticas públicas, exploração de efeitos de fatores determinantes em variáveis socioeconômicas e, além disso, trabalhos voltados à busca de soluções ou aprimoramento de instrumentos de gestão e de auditoria pública. A publicação traz, ainda, análises reflexivas, resultados de laboratórios conduzidos em diferentes áreas de políticas públicas e exemplares de meios de disseminação de conhecimento que exploraram recursos gráficos e de síntese de evidências também adquiridos em curso.

Trata-se, portanto, de uma contribuição que interessará a analistas e a servidores atuantes nas áreas de políticas pesquisadas, aos estudiosos do campo e aos gestores de organizações públicas e áreas de desenvolvimento de pessoas que busquem capacitar suas equipes nesse campo de conhecimento que merece continuar a ser apropriado e estudado como meio para a melhoria do serviço público.

REFERÊNCIAS

AMARA, N., OUIOMET, M., LANDRY, R. **New Evidence on Instrumental, Conceptual, and Symbolic Utilization of University Research in Government Agencies.** *Science Communication* 26, 75-106, 2004. Disponível em: <<https://doi.org/10.1177/1075547004267491>>. Acesso em: 10 out. 2022.

DONOHO, D. **50 Years of Data Science.** *Journal of Computational and Graphical Statistics* 26, 745-766, 2017. Disponível em: <<https://doi.org/10.1080/10618600.2017.1384734>>. Acesso em: 10 out. 2022.

FILGUEIRAS, F., LUI, L. **Designing data governance in Brazil:** an institutional analysis. *Policy Design and Practice* 1-16, 2022. Disponível em: <<https://doi.org/10.1080/25741292.2022.2065065>>. Acesso em: 10 out. 2022.

FISCHER, F.; GOTTWEIS, H. **The argumentative turn revisited:** public policy as communicative practice. Durham; London: Duke University Press, 2012.

FURTADO, B., LASSANCE, A. Simulações computacionais aplicadas à tomada de decisão pública, in: **Políticas públicas e usos de evidências no Brasil:** conceitos, métodos, contextos e práticas. Ipea, Brasília, pp. 195-222. 2022.

KITCHIN, R. **Big Data, new epistemologies and paradigm shifts.** *Big Data & Society* 1, 205395171452848, 2014. Disponível em: <<https://doi.org/10.1177/2053951714528481>>. Acesso em: 10 out. 2022.

KOGA, N., PALOTTI, P., MELLO, J., PINHEIRO, M.M.S. **Políticas públicas e usos de evidências no Brasil:** conceitos, métodos, contextos e práticas. Ipea, Brasília, 2022.

LEJANO, R. P. **Frameworks for policy analysis:** merging text and context. Routledge, New York, 2006.

LINDBLOM, E. **The science of muddling through public administration.** *Public Administration Review*, v. 19, n. 2, p. 79-88, 1959.

OUIOMET, M., LANDRY, R., ZIAM, S., BÉDARD, P.-O. **The absorption of research knowledge by public civil servants.** *Evidence & Policy: A Journal of*

Research, Debate and Practice 5, 331-350, 2009. Disponível em: <<https://doi.org/10.1332/174426409X478734>>. Acesso em: 10 out. 2022.

SIMON, H. A. **Rational choice and the structure of the environment**. Psychological Review 63, 129-138, 1956. Disponível em: <<https://doi.org/10.1037/h0042769>>. Acesso em: 10 out. 2022.

PINK, P. **Beyond public policy: a public action languages approach**. Northampton: Edward Elgar Publishing, 2019.

INTRODUÇÃO

ALEX LOPES PEREIRA ⁽³⁾

HÉLIO BOMFIM DE MACÊDO FILHO ⁽⁴⁾

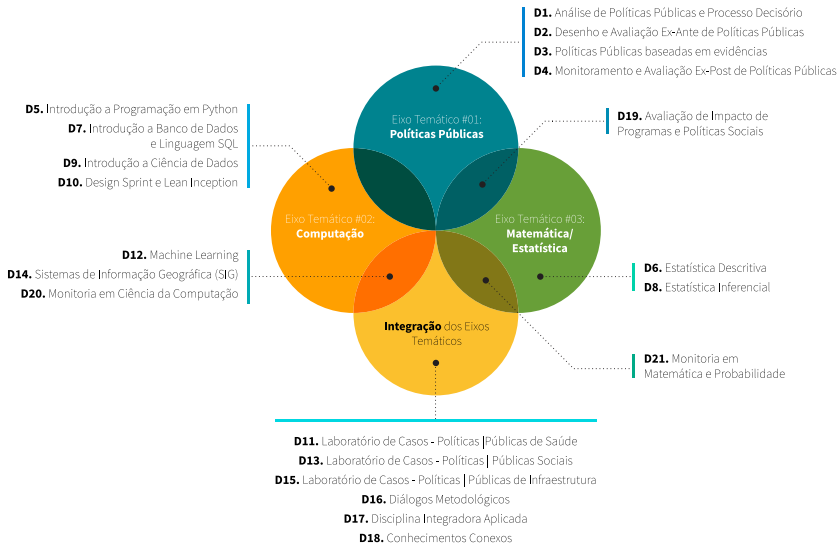
A Escola Nacional de Administração Pública (Enap), em parceria com a Controladoria-Geral da União (CGU), promove a Especialização em Ciência de Dados Aplicada a Políticas Públicas. Nessa iniciativa inovadora, servidores e empregados públicos federais demonstram, em Laboratórios de Casos e Trabalhos de Conclusão de Curso, contribuições relevantes no campo de Políticas Públicas Baseadas em Evidências. Na ocasião de sua segunda edição, a Especialização capacitou trinta e cinco agentes públicos. Esta publicação é um retrato dos resultados da segunda edição desta iniciativa pioneira.

Em seu projeto pedagógico, a Especialização está estruturada nos eixos temáticos de Computação, Matemática/Estatística e Políticas Públicas. Ciclos de palestras e Laboratórios de casos entrelaçam os conhecimentos dos eixos temáticos ao longo da especialização.

O processo de capacitação da segunda edição da Especialização ocorreu entre 5 de abril de 2021, data do início das aulas, e 14 de setembro de 2022, data da entrega dos Trabalhos de Conclusão de Curso. O resultado desta iniciativa é concretizado nas avaliações formativas e somativas.

Nas avaliações formativas, com as disciplinas de Laboratório de Casos, os participantes, em grupos, aplicam os conhecimentos adquiridos circunscritos aos contextos de Políticas Públicas de Saúde, Sociais e de Infraestrutura, com bases de dados prospectadas pelos facilitadores.

Figura 1 | Síntese da Programação da Especialização em Ciência de Dados Aplicada a Políticas Públicas da Escola Nacional de Administração Pública



Nas avaliações somativas, com os Trabalhos de Conclusão de Curso, os participantes, individualmente, aplicam os conhecimentos adquiridos nas distintas Políticas Públicas brasileiras.

As avaliações formativas e somativas reuniram aplicações reais de estatística descritiva, estatística inferencial, aprendizado de máquina supervisionado, aprendizado de máquina não supervisionado, processamento de linguagem natural, entre outras técnicas de ciência de dados, em temas de relevância nacional.

A pluralidade de temas abordados, bem como a extensão das técnicas aplicadas, foram razões determinantes para o desenvolvimento de uma publicação que desmistifique e dissemine a Ciência de Dados como campo multidisciplinar do conhecimento no cenário brasileiro de Políticas Públicas. Com efeito, tem-se a expectativa de que os leitores desta obra sejam inspirados pelas peças aqui presentes.

Dito isso, esta publicação contém uma seleção de trabalhos da segunda edição da Especialização em Ciência de Dados aplicada a Políticas Públicas da Escola Nacional de Administração Pública, que buscou o equilíbrio entre

a relevância do tema proposto, bem como a variedade e a robustez de aplicações do ferramental de Ciência de Dados.

Dos Laboratórios de Casos, dois estão presentes nesta publicação, sendo ambos de Políticas Públicas Sociais, cobrindo temas como cadastro único para programas sociais, bem como segurança alimentar e nutricional. As técnicas incluem aplicações de Estatística Descritiva e Aprendizado de Máquina Supervisionado.

Dos Trabalhos de Conclusão de Curso, nove estão presentes nesta publicação, cobrindo temas como Auditoria, Bancos Públicos, Compras Governamentais de Tecnologia da Informação, Desmatamento Ilegal da Amazônia, Ensino Superior, Meio Ambiente e Produção de Biocombustível, Regulação da Inteligência Artificial, Sistema de Recomendação para Categorização de Compras Governamentais e Sistema Único de Saúde. As técnicas incluem aplicações de Aprendizado de Máquina Não Supervisionado, Aprendizado de Máquina Supervisionado, Detecção de Anomalia, Estatística Descritiva e Estatística Inferencial, Mineração de Texto com Vetorização, Pesquisa Normativa, Processamento de Linguagem Natural e Sistemas de Informação Geográficos.

Em breve síntese, esta obra é o resultado de um esforço coletivo de docentes, discentes e profissionais de apoio, das mais variadas formações e com conhecimento em diversos campos de atuação. Há que se destacar, sobretudo, os discentes, capazes de traduzir as teorias aprendidas durante o curso na prática, materializadas nas peças aqui selecionadas.

Quadro 1 | Panorama de Políticas Públicas, Temas e Ferramental de Ciência de Dados Explorados na Publicação

	Política Pública (Área)	Tema	Ferramenta da Ciência de Dados
Laboratório de Casos	Social	Cadastro Único para Programas Sociais	Estatística Descritiva
	Social	Segurança Alimentar e Nutricional	Aprendizado de Máquina Supervisionado

	Política Pública (Área)	Tema	Ferramenta da Ciência de Dados
Trabalho de Conclusão de Curso	Gestão Pública	Compras Governamentais de Tecnologia da Informação	Aprendizado de Máquina Supervisionado
	Regulatória	Regulação da Inteligência Artificial	Pesquisa Normativa
	Educacional	Ensino Superior	Aprendizado de Máquina Supervisionado
	Conformidade e Controle	Auditoria	Aprendizado de Máquina Não Supervisionado
	Econômica	Bancos Públicos	Estatística Inferencial Aprendizado de Máquina Supervisionado
	Gestão Pública	Sistema de Recomendação para Categorização de Compras Governamentais	Mineração de Texto com Vetorização Aprendizado de Máquina Supervisionado
	Responsabilização (Accountability) e Ambiental	Desmatamento Ilegal da Amazônia	Estatística Descritiva
	Energética e Ambiental	Meio Ambiente e Produção de Biocombustível	Estatística Descritiva
	Saúde	Sistema Único de Saúde	Processamento de Linguagem Natural Detecção de Anomalia

Pretende-se que o livro sirva como referencial para as próximas edições da Especialização, além de ser uma iniciativa pioneira a abrir mais espaço para um serviço público de qualidade com a incorporação de novas tecnologias projetadas para o ganho de escala e, principalmente, para o aperfeiçoamento do dia a dia da Administração Pública, especialmente no cenário nacional de Políticas Públicas.

Ainda que aqui esteja disponível um amplo espectro de aplicações da Ciência de Dados nas Políticas Públicas brasileiras, sabemos que há um longo caminho a ser trilhado no futuro. Dadas as considerações iniciais, convidamos, doravante, o leitor a desfrutar do material a seguir. Bom proveito!



**TEMA: COMPRAS
GOVERNAMENTAIS
DE TECNOLOGIA DA
INFORMAÇÃO**

FERRAMENTA: APRENDIZADO DE
MÁQUINA SUPERVISIONADO

ANÁLISE DESCRITIVA DAS DESPESAS DE TECNOLOGIA DA INFORMAÇÃO DOS ÓRGÃOS DO PODER EXECUTIVO FEDERAL¹

ALEXANDRE MARTINS NOGUEIRA ⁽⁵⁾

INTRODUÇÃO

No contexto do planejamento das auditorias da Coordenação-Geral de Auditoria de Tecnologia da Informação (CGATI) junto aos órgãos e às entidades do Governo Federal do Brasil, um conjunto de dados bastante útil é a execução orçamentária e financeira das despesas de Tecnologia da Informação (TI). Isso porque as análises perpetradas sobre esses dados permitem identificar objetos de maior materialidade, considerando o volume financeiro dessas despesas, bem como itens de grande relevância na contratação das empresas públicas de TI (Serpro e Dataprev), a exemplo das despesas voltadas para a contratação de serviços de hospedagem de sistemas.

1 Este artigo resulta de trabalho de conclusão orientado pelo Professor Gustavo Cordeiro Galvão Van Erven.

Em uma outra dimensão, um planejamento baseado nesses dados permite que as unidades de auditoria reorganizem sua estrutura, desenvolvam novas competências técnicas para que, redefinindo suas prioridades, possam atuar de maneira mais eficiente. A problemática tratada nesta pesquisa diz respeito à existência de lacunas informacionais, cuja causa principal reside na ausência de ferramentas intuitivas e amigáveis, a exemplo de painéis gerenciais, dentro e fora da Controladoria-Geral da União, que trazem não somente uma visão amigável dos dados, mas também a extração do dado em si, no seu detalhe.

Com efeito, o objetivo geral deste trabalho foi a elaboração de um modelo de análise descritiva das despesas de TI dos órgãos do PEF (Poder Executivo Federal), a partir de dados extraídos do Sistema Integrado de Administração Financeira do Governo Federal (SIAFI), com a finalidade de incrementar o planejamento das auditorias de TI da CGATI. Segundo Soares (2020, p. 10):

A ciência de dados aplicada à auditoria interna permite analisar maior quantidade de informação em maior frequência, provendo auditorias mais abrangentes e em menor espaço de tempo. Com a utilização de maior quantidade de dados, o processo de previsão dos riscos se torna mais preciso e possibilita automatizar processos repetitivos, deixando-os mais ágeis e liberando os auditores de processos manuais que consomem muita força de trabalho.

Os resultados demonstraram o êxito do modelo ao utilizar como critério de seleção dos empenhos de TI sua classificação em nível de natureza da despesa detalhada, que contém, em sua estrutura básica de codificação, o elemento de despesa, ou seja, o objeto do gasto. Porém, conclui-se que se faz necessário cruzar os dados do SIAFI com outras bases oficiais, tais como o Sistema Integrado de Administração de Serviços Gerais (SIASG), no intuito de refinar a identificação do objeto do gasto de TI.

Com relação à descrição dos dados, as análises exploratória e explanatória sobre os dados transformados (obtidos após limpeza, formatação e adição de novas colunas) permitiram demonstrar principalmente que, em que pese o notório aumento dos gastos de TI no período considerado (2017 a 2021), cerca de 80% desse volume refere-se ao custeio da infraestrutura de TI, tendo sido realizados poucos investimentos (despesas de capital). Outro dado relevante diz respeito à proporção dos gastos de TI em relação ao total de gastos, que alcançou o índice médio de 6% no exercício de 2021, patamar próximo ao identificado no setor privado.

No decorrer do período considerado, verificou-se significativo aumento nas despesas de TI dos órgãos do Poder Executivo Federal junto às empresas públicas de TI, principalmente, Serpro e Dataprev, podendo indicar uma preferência do governo federal em manter total controle de sua plataforma tecnológica, considerando os riscos crescentes de vazamentos de dados. Com relação aos objetos de TI, verificou-se que há predominância de compras de equipamentos de microinformática e contratação de serviços de desenvolvimento e sustentação de *software* e de infraestrutura de TI (*service desk* e suporte à infraestrutura tecnológica do órgão).

DESENVOLVIMENTO

O PROCESSO DE TRABALHO DESTA PESQUISA OBEDECEU ÀS SEGUINTE ETAPAS:

- a) Coleta e preparação dos dados — consistiu na extração dos empenhos emitidos no período, 2017 a 2021, a partir de consulta construída no Tesouro Gerencial (TG, cuja fonte de dados é o SIAFI), filtrando as despesas de TI pela sua natureza da despesa detalhada (NDD). Inicialmente, foi feito um levantamento cuidadoso para identificar os 111 códigos de NDD que seriam utilizados na filtragem dos empenhos, a saber:

44903793; 44903993; 44913993; 44903047; 33903047; 44904005; 44905255; 44904006; 33903994; 33913994; 44903646; 33913047; 33904017; 33904018; 33904019; 33914019; 33903997; 33913997; 449039

97;33904013;33914013;33903504;33913504;44903504;33904008
;44903645;44904001;33903926;33913926;44903992;33904015;3
3904023;33914023;44915235;44905237;44905241;44905245;449
05243;44905247;33903930;33913930;33904009;33803947;33904
025;44904009;33913931;33903931;33904001;33914001;3390400
2;33904004;33904003;33904005;33903911;33904006;33913911;
33903995;33913995;33914007;33904007;33903709;33903908;33
913908;44903792;44904002;33903017;33913017;44903017;4490
5235;44903995;44903654;44904004;44904010;33904026;339040
99;33904016;33913957;33903956;33913956;33909240;33919240
;44909240;33904024;44904007;33904096;44904096;44903657;3
3903655;33903957;44903957;33903657;33904021;33914021;449
04003;33903728;33903928;33913928;33904010;33914010;33903
727;33903927;33904011;33914011;33913927;33903056;3391401
4;33904014;33904020;33914020;33904022;33914022.

- b) Preparação dos dados — consistiu no trabalho de *Modelagem dos dados*, envolvendo limpeza, formatação, transformação de colunas e adição de novas colunas, visando agregar mais dados ao modelo (ex.: o objeto de TI, conforme o universo de auditoria de TI da CGU; ou se o empenho tem como favorecido uma das estatais de TI). Foi desenvolvida a *Rotina ETL* (Extract-Transform-Load) na linguagem *Python*. Utilizou-se ainda modelos de aprendizagem de máquina supervisionado (baseados na *biblioteca scikit-learn*) para o preenchimento de colunas que identificam o objeto dos empenhos registrados em *naturezas da despesa genéricas*². O modelo de classificador utilizado para esse fim foi configurado da seguinte maneira:
- **Tipo de modelo ML utilizado:** classificador (variáveis categóricas).
Dados históricos: 2.347 registros (rotulagem ou anotação manual).
Total de categorias criadas: 9 (nove). Massa de treino: 79% dos registros (Teste: 21%). Massa de treino após o resample: 6.786 registros (754 por

2 Naturezas de despesa cuja definição não permite identificar precisamente o objeto de TI adquirido ou contratado. São eles: Serviços Técnicos Profissionais de TIC, Serviços de Tecnologia da Informação, Serviços de Tecnologia da Informação - PJ, Outros serviços de TIC, Serviços de TI - PJ de Caráter Secreto ou Reservado, Serviços de TI - PJ Pagamento Antecipado, Multas e Juros.

categoria), com SMOTEN (All). Hiperparâmetros: SGDClassifier(loss='log', penalty='elasticnet', alpha=1e-06, tol=0.0001). Algoritmo escolhido: SGDClassifier (scikit-learn), próprio para classificador multcategórico. Acurácia: em torno de 86%.

- c) Análise exploratória e explanatória — nessa etapa foi realizada a análise descritiva das despesas de TI dos órgãos e das entidades públicos federais (constituídos de 32 órgãos superiores, 233 órgãos e 2.904 unidades gestoras executoras). Foram elaborados resumos estatísticos (estatística descritiva, com medidas de tendência central e de dispersão) e gráficos diversos, adotando os princípios de Gestalt³ na visualização dos dados, utilizando cores para destacar as informações consideradas relevantes (“pérolas”) para o contexto de auditoria. De acordo com Reis, E. e Reis, I. (2002, p. 5):

Análise Descritiva é a fase inicial deste processo de estudo dos dados coletados. Utilizamos métodos de Estatística Descritiva para organizar, resumir e descrever os aspectos importantes de um conjunto de características observadas ou comparar tais características entre dois ou mais conjuntos. As ferramentas descritivas são os muitos tipos de gráficos e tabelas e também medidas de síntese como porcentagens, índices e médias.

A partir dessa análise exploratória, direcionada por perguntas sobre os dados (P1 a P9), foram feitas explicações em formato textual sobre as informações apresentadas em formato visual.

RESULTADOS DAS ANÁLISES EXPLORATÓRIA E EXPLANATÓRIA

P1 | O que as despesas de TI apresentam de relevante em seu resumo estatístico?

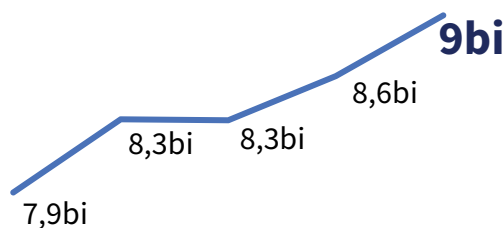
³ Os princípios da Gestalt e os estudos sobre percepção visual auxiliam no entendimento de como os indivíduos compreendem as informações transmitidas, sendo possível eliminar elementos desnecessários dos dashboard para facilitar o processamento da informação apresentada (KNAFLIC, 2019).

Extrato geral das principais variáveis qualitativas categóricas	Resumo estatístico do valor dos empenhos
Qtde. de registros: 290.688	Total empenhado:..... 42 bi
Qtde. de Órgãos Superiores:..... 33	Média:..... 145.288,21
Qtde. de Órgãos:..... 233	Mediana: 2.713,00
Qtde. Unidades Gestoras Executoras: 3.473	Desvio padrão:3.449.874,26
Qtde. de Favorecidos (pelo nome): ... 20.163	Máximo: 797.429.247,77
	Mínimo: 0,01
	Amplitude: 797.429.247,76
	Primeiro quartil(Q1): 676,51
	Segundo quartil (Q2): 2.713,00
	Terceiro quartil(Q3): 13.499,23

Totalizando R\$ 42 bi no período considerado, o conjunto de dados de empenhos de TI não é uniforme. Ele possui alta variabilidade de valores (com grande dispersão, conforme se observa no gráfico), em virtude do *desvio padrão* de valor elevado (R\$ 3.449.874,26) em relação à *média aritmética* (R\$ 145.288,21), bem como a elevada *amplitude* (R\$ 797.429.247,76). Há uma concentração de empenhos com pequenos valores, considerando que 50% dos empenhos (situados entre os *quartis Q1 e Q3*) situam-se entre R\$ 676,52 e R\$ 13.499,23, valores significativamente abaixo da média.

P2 | Qual o comportamento dos gastos de TI no período de 2017 a 2021?

De 2017 a 2021, observa-se um aumento gradativo das despesas de TI, tendo sido identificada leve queda de 2018 para 2019 (redução de 0,06%), provavelmente motivada por cortes do orçamento de 2019, além de outras medidas adotadas para evitar ultrapassar o teto de gastos. Nos anos seguintes, as despesas continuaram aumentando e, em 2021, as despesas de TI (empenhos emitidos) alcançaram o patamar de cerca de R\$ 9 bi, o que aparentemente representa um aumento de 13% em relação a 2017:

Gráfico 1 | Evolução das despesas de TI (2017 a 2021)

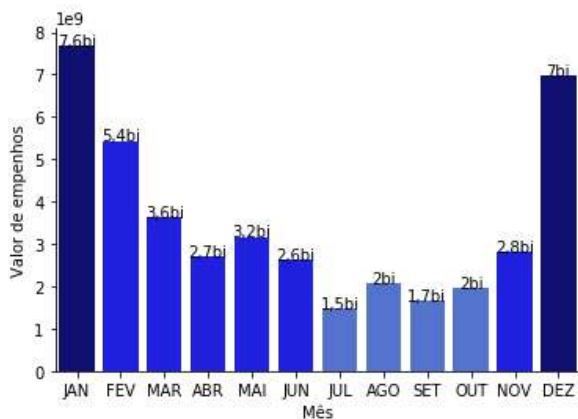
Ano	Valor (bi)
2017	7,9
2018	8,3
2019	8,3
2020	8,6
2021	9

Tabela 1 | Despesas de TI p/ exercício (2017 a 2021)

Exercício	Total dos empenhos (R\$)	Aumento %
2017	7.951.655.541,19	—
2018	8.364.264.255,56	5,19%
2019	8.359.338.639,37	-0,06%
2020	8.607.851.808,77	2,97%
2021	8.950.429.483,89	3,98%
Total	42.233.539.728,78	—

Entretanto, há uma ressalva importante: ao realizarmos a correção monetária do total empenhado em 2017 (R\$ 7.951.655.541,19 ou 7,9 bi), adotando o Índice de Custo de Tecnologia da Informação (ICTI)⁴, tem-se que o valor corrigido alcança um valor superior em relação a 2021: R\$ 9.675.347.720,68 (R\$ 9,7 bi). Isso indica que, na realidade, as despesas de TI sofreram redução de 7,5%, comparando-se 2017 com 2021.

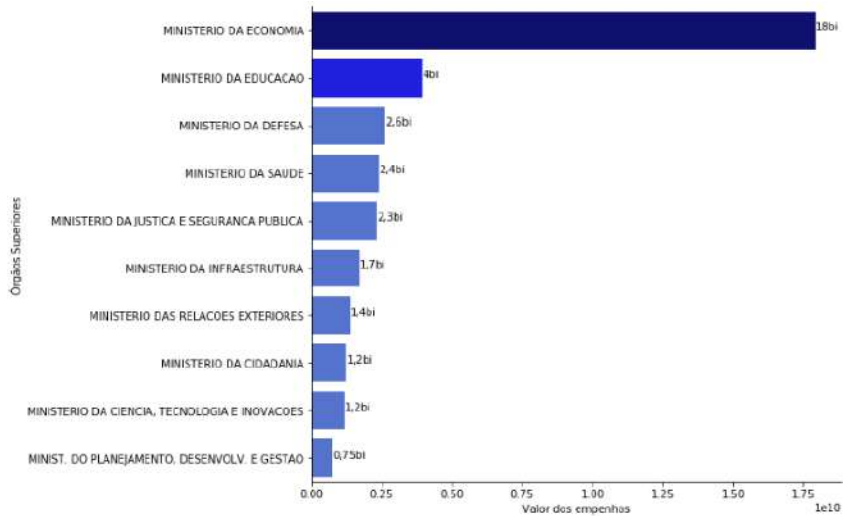
4 Série histórica do Índice de Custos da Tecnologia da Informação. IPEA. Disponível em: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.ipea.gov.br%2Fcartadeconjuntura%2Fwp-content%2Fuploads%2F2021%2F12%2F211203_icti_serie_completa_setembro21.xlsx&wdOrigin=BROWSELINK.

P3 | Qual o comportamento dos gastos de TI mês a mês?**Gráfico 2** | Evolução das despesas de TI por mês de emissão

Há uma concentração maior de empenhos emitidos no início do ano (janeiro e fevereiro) e no final do ano (dezembro). A hipótese considerada é a de que esse comportamento dos órgãos é motivado pela necessidade de evitar que o orçamento do órgão tenha uma execução baixa ao final do exercício, o que poderia resultar em uma redução no seu orçamento para o exercício seguinte. Outra hipótese é a de que, durante o exercício (de abril a setembro, no máximo até outubro), o orçamento geral sofre bloqueios, situação que apenas se normaliza a partir de novembro, com o desbloqueio desses recursos.

P4 | Quais os órgãos que concentram a emissão de empenhos de TI no período, em termos de volume financeiro?

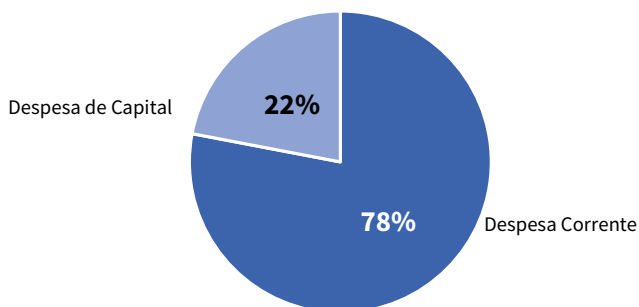
Gráfico 3 | Ranking dos 10 maiores consumidores de bens e serviços de TI (para órgão superior)



Ministério da Economia, Ministério da Educação e Ministério da Defesa são os órgãos superiores que concentram mais da metade das despesas de TI relativas ao período considerado (24,6 bi, ou 59% de 42 bi). O Ministério da Economia, que aglutinou as funções e as competências dos extintos Ministério da Fazenda e Ministério do Planejamento (desde 2019), é o órgão superior que concentra R\$ 18 bi em despesas de TI (43% do total de 42 bi). E mais: 81% do total de R\$ 18 bi refere-se às despesas de TI do próprio Ministério (14,5 bi), seguido do INSS, que concentra 10% desse total (10%).

Na área da educação, o gasto de TI encontra-se distribuído de maneira mais uniforme entre os cinco primeiros órgãos, estando na dianteira o próprio Ministério da Educação e o Inep (órgão que realiza e aplica a prova do Enem). Os Comandos da Aeronáutica e do Exército lideram no âmbito do Ministério da Defesa (respectivamente, com 43 e 40% do total de 2,6 bi).

P5 | Qual o percentual de empenhos emitidos referentes às despesas correntes e de capital, em termos financeiros?

Gráfico 4 | Despesas de TI por categoria econômica da despesa

Segundo o MTO 2021⁵, as despesas de capital contribuem, diretamente, para a formação ou a aquisição de um bem de capital, incluindo gastos para a produção ou para a geração de novos bens ou serviços que integrarão o patrimônio público, ou seja, representam acréscimos ao patrimônio da União. Na área de TI, são exemplos de despesas de capital a compra de equipamentos e materiais permanentes, inclusive a aquisição de licenças de *software*. Por outro lado, as despesas correntes representam todas as demais despesas, aquelas que não contribuem, diretamente, para a formação ou a aquisição de um bem de capital e envolvem gastos com a manutenção e o funcionamento dos serviços públicos em geral (gastos com custeio da infraestrutura). Na área de TI, são exemplos a contratação de serviços de sustentação de infraestrutura de TI, a sustentação de *software* e a hospedagem de sistemas.

P6 | Qual a proporção das despesas de TI (empenhos emitidos), em termos percentuais, em relação ao total de despesas de cada órgão?

Um aspecto considerado relevante quando analisamos as despesas dos órgãos e das entidades governamentais diz respeito à verificação de quanto o orçamento anual é dedicado às despesas com tecnologia da informação. Essa informação pode ser útil para inferir a maturidade organizacional relacionada à gestão de TI, ou ainda, se há o risco de que a infraestrutura de TI

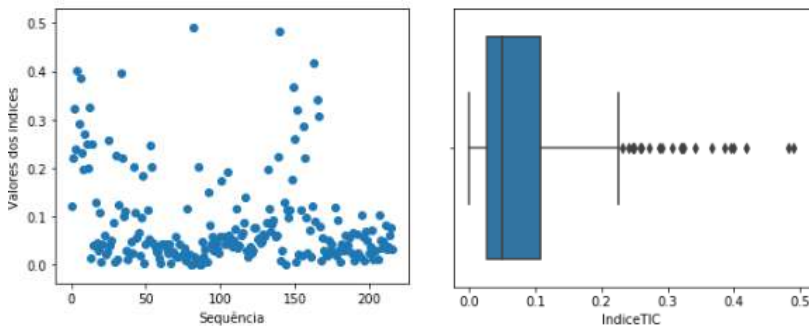
⁵ Manual Técnico do Orçamento relativo ao exercício de 2021. Disponível em: <https://www1.siof.planejamento.gov.br/mto/lib/exe/fetch.php/mto2021:mto2021-versao4.pdf>. Acesso em: 27 maio de 2022.

daquele órgão esteja em situação precária (sucateamento). Uma forma de obter essa informação é calcular, para cada exercício financeiro, a proporção das despesas de TI em relação ao total de despesas administrativas (extraída mediante consulta construída no TG), considerando TI e Não-TI. Adotando a abordagem dessa segunda alternativa, para efeitos das análises constantes desta pesquisa, foi construído o Índice de Proporção das Despesas de TI a partir da seguinte fórmula de cálculo:

$$\text{Índice TIC} = (\sum \text{Valor do empenho de TI}) / (\sum \text{Valor dos empenhos})$$

Após a aplicação dessa fórmula de cálculo para todos os 217 órgãos distintos, foi possível obter o índice individual para cada órgão e exercício financeiro. Para obter o índice médio de proporção de gastos de TI nos órgãos do PEF, focou-se no exercício de 2021, partindo da plotagem dos valores do referido índice em gráficos de dispersão e de quartis (*boxplot*) e a partir do resumo estatístico dos valores do índice:

Gráfico 5 | Gráfico de dispersão e boxplot do conjunto de valores do Índice de Proporção dos Gastos de TI do exercício de 2021



Os gráficos de dispersão e *boxplot* relativos à distribuição dos valores dos índices de 2021 demonstram que os dados são significativamente dispersos (média aritmética = 0.0867 ou cerca de 9%; desvio padrão = 0.0974), indicando a presença de valores discrepantes ou destoantes (*outliers*). Assim, mostra-se necessário sanitizar ou sanear os dados recalculando a média sem os valores

discrepantes utilizando o método clássico baseado na *amplitude interquartil*, cujos cálculos encontram-se reproduzidos a seguir:

Calculando o Índice Médio de Proporção de Gastos de TI sem outliers

Primeiro quartil(Q1): 0.0263

Segundo quartil (Q2): 0.0487

Terceiro quartil(Q3): 0.1072

Método para identificação de outliers - Amplitude Interquartil (DQ):

$$DQ = Q3 - Q1 = 0.0808$$

$$\text{Limite Superior} = Q3 + ((1.5) \times DQ) = 0.2284$$

$$\text{Limite Inferior} = Q3 - ((1.5) \times DQ) = -0.0948$$

Após a eliminação de 22 outliers (sanitização do conjunto), a média aritmética passou de 0.0867 (9%) para 0.0599 (6%).

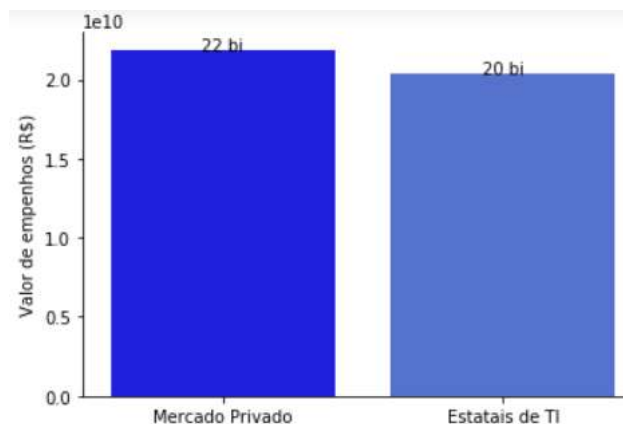
Após a sanitização dos dados, verificou-se que o índice médio de proporção dos gastos de TI alcança o valor aproximado de 0,0599 (ou 0,06), significando que as despesas de TI nos 217 órgãos do PEF analisados representam, em média, 6% do total das despesas do exercício de 2021. Esse patamar está muito próximo ao identificado pela Fundação Getúlio Vargas, em estudo publicado em 2018⁶, que aponta uma relação de 7,7% das despesas de TI em relação à receita de médias e grandes empresas nacionais de capital privado.

P7 | Qual o total empenhado tendo como favorecido apenas as estatais de TI? E com relação ao mercado privado?

A seguir, será apresentado um comparativo entre os gastos de TI junto ao mercado privado e os gastos de TI tendo como favorecidas as principais estatais (dependentes ou não) que reconhecidamente prestam serviços na área de TI (Serpro, Dataprev, Telebrás, Caixa Econômica Federal e Banco do Brasil):

6 29ª Pesquisa Anual do Uso de TI, 2018. FGV. Disponível em: <https://eaesp.fgv.br/sites/eaesp.fgv.br/files/pesti2018gvciappt.pdf>.

Gráfico 6 | Volume financeiro de empenhos emitidos de 2017 a 2021
(Estatual x Mercado Privado)



Mercado Privado	Estatais de TI
Qtde. empenhos:..... 281.258	Qtde. empenhos:.....9.430
Valores empenhados:.....21.881.919.518,83	Valores empenhados:.....20.351.620.209,95
Média:..... 77.800,17	Média:..... 2.158.178,18
Mediana:2.741,35	Mediana:1.532,00
Desvio padrão:..... 666.703,16	Desvio padrão:..... 18.694.092,85

Da média dos R\$ 42 bi empenhados no período, R\$ 20 bi (ou 48% do total) referem-se a empenhos de TI emitidos em favor das estatais de TI considerados neste estudo (Serpro, Dataprev, Telebrás, Banco do Brasil e Caixa Econômica Federal). Percebe-se, pelos resumos estatísticos, que ambos os conjuntos de dados são dispersos (com desvio padrão significativamente superior à média), indicando a presença de *outliers*.

Importante mencionar que, comparando-se os anos de 2017 e 2021 de maneira isolada, verifica-se um aumento significativo de mais de 32% nas despesas de TI dos órgãos do PEF junto a essas instituições, podendo indicar uma preferência do governo federal em manter total controle de sua plataforma tecnológica, provavelmente em virtude dos riscos crescentes de vazamentos

de dados e de critérios rígidos dispostos na Lei Geral de Proteção de Dados (LGPD) e regulados pela Autoridade Nacional de Proteção de Dados (ANPD):

Tabela 2 | Comparativo das despesas de TI com estatais de TI (Comparativo 2017 versus 2021)

Total dos empenhos em 2017 (R\$)	Total dos empenhos em 2021(R\$)	Aumento (%)
3.589.080.233,03	4.743.488.047,25	32,16%

P8 | Quais favorecidos prestam serviços para os órgãos e para as entidades públicos federais e qual a participação de mercado de cada uma deles?

Há predomínio da estatal Serpro (73% do total de 20,3 bi) constando como favorecida dos empenhos de TI emitidos no período de 2017 a 2021. Já as contratações de serviços de TI junto às empresas do mercado privado alcançam o montante de R\$ 15,3 bi (70% do total de 22 bi), volume expressivo em comparação à aquisição de bens de TI (30%), conforme a tabela a seguir:

Tabela 3 | Participação de mercado dos principais favorecidos dos empenhos de TI no período de 2017 a 2021 (Estatal x Mercado privado)

Grupo 1 – Empresas públicas de TI e outras estatais		
Nome da Estatal	Valor dos empenhos(R\$)	Participação%
SERPRO	14.767.249.759,17	73%
DATAPREV	4.754.265.751,87	23%
TELEBRAS	799.488.477,94	4%
CONSÓRCIO SERPRO/DATAPREV	22.247.505,46	0,1%
BANCO DO BRASIL	7.445.208,60	0,04%
CAIXA ECONÔMICA FEDERAL	923.506,91	0,005%
TOTAL	20.351.620.209,95	—

Grupo 2 – Mercado privado

Nome do Favorecido pertencente ao mercado privado (AQUISIÇÃO DE BENS DE TI)	Total dos empenhos (R\$)	Part. %
DELL COMPUTADORES DO BRASIL LTDA	700.365.026,64	11%
POSITIVO TECNOLOGIA S.A.	445.331.065,85	7%
CONSÓRCIO TEPRO	295.029.720,53	4%
TORINO INFORMATICA LTDA.	269.962.362,08	4%
SYSTECH SISTEMAS E TECNOLOGIA EM INFORMATICA LTDA	198.371.987,98	3%
OUTROS (13.645 empresas distintas)	4.651.119.754,50	71%
Total	6.560.179.917,58	—

Nome do Favorecido (SERVIÇOS DE TI)	Total dos empenhos (R\$)	%
CLARO S.A.	968.423.051,28	6%
CENTRAL IT TECNOLOGIA DA INFORMACAO LTDA	689.640.927,83	5%
CAST INFORMATICA S/A	553.620.952,51	4%
BRASOFTWARE INFORMATICA LTDA	452.188.654,29	3%
CTIS TECNOLOGIA LTDA	450.113.199,02	3%
OUTROS (8.714 empresas distintas)	12.207.751.509,91	80%
Total	15.321.738.294,84	—

De acordo com os dados, no que se refere às despesas com serviços de TI relativos ao mercado privado, verifica-se que a participação de mercado é bastante pulverizada, ou seja, não há concentração evidente nas mãos de apenas uma empresa ou grupo de empresas. Entretanto, destacam-se nesse segmento as despesas com a CLARO S.A., a CENTRAL IT e a CAST.

Quanto às despesas com aquisição de bens de TI relativas ao mercado privado (também pulverizado), embora a DELL apresente-se como líder nesse ramo (com uma fatia de 11%), não se evidenciam grandes concentrações em apenas uma empresa ou um grupo de empresas.

P9 | Quais os principais objetos de TI adquiridos ou contratados pelos órgãos e pelas entidades públicos federais no período considerado?

Tabela 4 | Despesas de TI segmentado por objeto de TI (Estatal x Mercado privado)

Grupo 1 - Empresas públicas de TI e outras estatais		
Objeto de TI	Total (R\$)	%
Hospedagem de Sistemas	7.428.050.277,88	36%
Sustentação de Infraestrutura de TI	5.457.609.269,54	27%
Desenvolvimento e Sustentação de Software	3.917.301.583,90	19%
Comunicação de Dados/ Internet	1.067.330.220,35	5%
Cloud Service (nuvem)	227.986.197,06	1%
Certificado Digital (serviço)	103.574.062,87	1%
Consultoria, treinamento e capacitação em TIC	75.641.644,29	0,4%
Software (compra ou locação)	43.060.417,05	0,2%
Outsourcing de impressão/ Digitalização de documentos	11.211.820,62	0,1%
Equipamentos de TI (locação)	2.732.802,77	0,01%
Material de TI (Bens)	2.177.166,26	0,01%
Equipamentos de TI (compra)	865,63	0,000004%
Outros serviços de TI	2.014.943.881,73	10%
TOTAL:	20.351.620.209,95	100%

Grupo 2 - Mercado privado

Objeto de TI	Total (R\$)	%
Material de TI (Bens)	3.800.836.562,81	17%
Desenvolvimento e Sustentação de Software	3.629.703.688,93	17%
Sustentação de Infraestrutura de TI	3.173.755.353,30	15%
Equipamentos de TI (compra)	2.759.343.354,77	13%
Software (compra ou locação)	2.727.260.575,35	12%
Comunicação de Dados/ Internet	2.358.664.908,79	11%
Equipamentos de TI (locação)	450.935.020,59	2%
Outsourcing de impressão/ Digitalização de documentos	434.557.905,94	2%
Consultoria, treinamento e capacitação em TIC	261.712.520,49	1%
Cloud Service (nuvem)	252.638.964,61	1%
Hospedagem de Sistemas	47.198.270,21	0,2%
Certificado Digital (serviço)	23.847.629,55	0,1%
Outros serviços de TI	1.961.464.763,49	9%
TOTAL	21.881.919.518,83	100%

No primeiro grupo, os dados mostram que os serviços de Hospedagem de Sistemas têm grande relevância dentro do contexto considerado, representando 36% (ou R\$ 7,4 bi) do volume total de despesas desse grupo (inclui *datacenter hosting* e *colcation, storage* em nuvem, administração e gestão de contas de e-mail, hospedagem e manutenção de sítios na internet). O serviço de operação e sustentação de infraestrutura de TI aparece em segundo lugar, com uma fatia de 27% (ou R\$ 5,4 bi) de participação em relação ao total; e, em terceiro, os serviços de desenvolvimento, manutenção e sustentação de

software. Os três primeiros serviços desse *ranking* representam juntos cerca de 80% do volume total de empenhos emitidos em favor das estatais de TI no período considerado (2017 a 2021).

No segundo grupo, os dados mostram que 7 (sete) objetos de TI concentram praticamente as aquisições de bens e contratações serviços de TI junto ao mercado privado (92%), com destaque para Material de TI (17% do total de R\$ 22 bi) e os serviços de Desenvolvimento e Sustentação de *Software* (16%) e Sustentação de Infraestrutura de TI (14%).

CONCLUSÃO

Esta pesquisa permitiu elaborar um *Modelo de análise e diagnóstico das despesas de TI* dos órgãos do Poder Executivo Federal (PEF) com base nos dados de execução orçamentária e financeira do SIAFI, sendo que o objetivo principal foi atingido: coletar e analisar as despesas de TI, contornando, de maneira documentada, os obstáculos e dificuldades técnicas encontrados, bem como estabelecendo uma linha de base de referência para a repetição sistemática do trabalho e evolução desta pesquisa em outras oportunidades.

Foram identificadas as plataformas disponíveis para acesso e extração dos dados de execução orçamentária e financeira dos órgãos do PEF, sendo a principal o Tesouro Gerencial (TG), ferramenta que extrai dados gerenciais do SIAFI, sistema que por sua vez realiza o registro e a guarda dos empenhos a partir da execução financeira do orçamento da União. Dentre as alternativas para extrair as despesas de TI, optou-se por escolher os empenhos na fase inicial de emissão, selecionando-os com base na natureza de despesa detalhada de TI. A seleção dos códigos de natureza de despesa de TI não se limitou apenas ao contido no Manual Técnico do Orçamento (MTO), mas foram identificados a partir da descrição das naturezas de despesa dos códigos extraídos diretamente no TG.

Apesar da natureza de despesa permitir identificar, em alguma medida, o objeto de TI que está sendo contratado ou adquirido, realizou-se o refinamento de alguns objetos mediante o uso de classificadores baseados em modelos

de aprendizado de máquina supervisionado aplicados sobre o campo “descrição” dos empenhos. Apesar de alguns registros não trazerem informações relevantes nesse campo, essa estratégia mostrou-se promissora, pois permitiu enquadrar um volume maior de empenhos dentro dos principais objetos de TI. Também foi possível prospectar os modelos de gráficos que ofereciam perspectiva razoável acerca das despesas de TI, sendo úteis ao desenvolvimento posterior de um protótipo de painel de gastos de TI (trabalho futuro).

Por fim, verificou-se que, a partir das despesas de TI, é possível obter informações úteis, relevantes e pertinentes para subsidiar o processo de planejamento de auditoria baseado em riscos. As análises permitiram identificar indicadores (ex.: Indicador de Proporção de Gastos de TI), extrair *insights* úteis (a exemplo de causas hipotéticas para certos comportamentos dos gastos) e outras informações relevantes para a auditoria, principalmente em termos de materialidade, criticidade e relevância.

Com relação ao diagnóstico sobre as despesas de TI dos órgãos e das entidades públicos federais relativo ao período de 2017 a 2021, verificou-se principalmente que:

1. No período considerado, foram emitidos cerca de R\$ 42 bi em empenhos de TI, sendo que 48% desse total refere-se a empenhos emitidos em favor das estatais de TI (Serpro, Dataprev, Telebrás, Banco do Brasil e Caixa Econômica Federal). Cerca de 80% desse volume de gasto destina-se ao custeio da infraestrutura de TI existente (despesa corrente).
2. Os gastos de TI aumentaram significativamente no período considerado. A emissão de empenhos aumentou sistematicamente nos últimos três meses do ano.
3. Ministério da Economia, Ministério da Educação e Ministério da Defesa são os órgãos superiores que concentram mais da metade das despesas de TI relativas ao período considerado.
4. O índice médio de proporção dos gastos de TI, para o exercício de 2021, alcançou o valor aproximado 6% do total de despesas, patamar próximo ao identificado no setor privado em estudo feito em 2018 pela FGV (7,7%).

5. Com relação aos empenhos de TI emitidos em favor das estatais de TI: foram emitidos o total de R\$ 20 bi em empenhos; comparando-se 2017 com 2021, verifica-se um aumento significativo de 32% do volume financeiro empenhado; o Serpro é a estatal de TI que lidera o *ranking* de participação do mercado (73% do volume financeiro), seguido de Dataprev e Telebrás.
6. Com relação aos empenhos de TI emitidos em favor de empresas do setor privado: foram emitidos o total de R\$ 22 bi em empenhos; comparando-se 2017 com 2021, houve pequena redução (de cerca de 4%) no volume financeiro empenhado; no setor de serviços, CLARO S.A., CENTRAL IT e CAST lideram a participação de mercado no período considerado; DELL e POSITIVO são os líderes de mercado no setor de compras de TI (bens); e, em termos financeiros, predomina a compra de equipamentos de TI, seguido da contratação de serviços de desenvolvimento e sustentação de *software* (ou “fábrica de *software*”) e serviços de operação de sustentação de Infraestrutura de TI.

Para trabalhos futuros, sugere-se principalmente: elaborar novas consultas construídas a partir do TG, objetivando a coleta de empenhos situados no estágio da liquidação e/ou do pagamento (conferindo maior acurácia ao levantamento das despesas de TI); refinar a seleção dos itens comprados ou contratados a partir de cruzamentos com o SIASG; identificar objetos de TI que se caracterizam pela alta concentração de mercado; aperfeiçoar o classificador baseado em algoritmo de aprendizado de máquina supervisionado adotado na etapa de pré-processamento dos dados, por exemplo, rotulando/ anotando um maior quantitativo de dados históricos; aperfeiçoar as análises trazendo mais visões sobre os dados, a exemplo de séries históricas dos gastos por órgão ou por tipo de objeto do gasto; desenvolver um Índice de Risco dos órgãos do PEF baseado no diagnóstico das despesas de TI; e ainda elaborar painel gerencial de gastos e de riscos de TI.

REFERÊNCIAS

KNAFLIC, C. N. **Storytelling com dados**: Um guia sobre visualização de dados para profissionais de negócios. Edição Colorida. Rio de Janeiro: Alta Books, 2019.

REIS, E.A.; REIS I.A. (2002). **Análise Descritiva de Dados. Relatório Técnico do Departamento de Estatística da UFMG**. Disponível em: <http://www.est.ufmg.br/portal/arquivos/rts/rte0202.pdf>. Acesso em: 10 maio 2022.

SOARES, G. F. Ciência de dados aplicada à Auditoria Interna. **Revista da CGU**, Vol. 12, Nº 22, Jul-Dez 2020. Disponível em: https://ojs.cgu.gov.br/index.php/Revista_da_CGU/article/view/195. Acesso em: 30 maio 2022.



**TEMA: REGULAÇÃO DA
INTELIGÊNCIA ARTIFICIAL**

FERRAMENTA:
PESQUISA NORMATIVA

PESQUISA QUALITATIVA E DESCRITIVA SOBRE BASE NORMATIVA NACIONAL DE IA À LUZ DE EXPERIÊNCIAS INTERNACIONAIS VOLTADAS AO USO DESSA TECNOLOGIA PELO PODER PÚBLICO¹

GUSTAVO DE QUEIROZ CHAVES⁽⁶⁾

INTRODUÇÃO

A tomada de decisão, no âmbito de políticas públicas, pode ser compreendida como sendo subsidiada pelas perspectivas de uma escolha de determinada agenda política somada aos insumos e às evidências trazidos pelas instâncias técnico burocráticas. Porém, independentemente da perspectiva considerada para justificar as decisões em políticas públicas, num Estado Democrático de Direito, essas precisam ser orientadas em respeito aos direitos fundamentais

1 Este artigo resulta de trabalho de conclusão orientado pelo Professor Alex Lopes Pereira.

garantidos pelo Estado legal e a outros valores comuns compartilhados e perseguidos por sua sociedade.

Inicialmente, faz-se necessária uma rápida definição de alguns termos relevantes para a compreensão das discussões acerca do que seria um sistema de Inteligência Artificial (IA), objeto quase central do presente trabalho. Em que pese não haver uma definição única para o que seria Inteligência Artificial, dentre algumas disponíveis na literatura, a Estratégia Brasileira de Inteligência Artificial (EBIA) adotou a definição da OCDE de que seria:

um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais. Os sistemas de IA são projetados para operar com vários níveis de autonomia (MCTI, 2021).

Essa definição incorpora um conceito muito relevante para as discussões: a forma como o resultado da IA impacta na entrega ao usuário, associado ao termo autonomia ao nível de automação. Em que pese, a depender do contexto de uso da IA haver diferentes formas de descrever seus níveis de autonomia, o trabalho utilizará a forma mais simples de classificação em três tipos, também utilizada pelo GAO (GAO, 2021): *man-in-the-loop*², *man-on-the-loop*³ e *man out of the loop*⁴.

A grande vantagem gerada pela alta velocidade de processamento de dados e de tomada de decisões pelos sistemas de IA, dentro de um contexto de globalização da informação, ao mesmo tempo que gera ganhos de eficiência e potencialmente melhores resultados na decisão, traz desafios de discussão

2 Processo decisório definido quando o ser humano possui o controle total sendo o revisor e a IA fornece apenas recomendações ou insumos para auxiliar em sua decisão.

3 Processo decisório no qual o ser humano está em uma função de monitoramento ou supervisão, com a capacidade de assumir o controle quando o modelo de IA encontrar eventos inesperados ou gerar resultados indesejáveis.

4 Processo decisório sem supervisão humana no qual a execução de decisões ocorre diretamente pelo sistema de IA concedendo-o o controle total sem a opção de anulação humana da decisão.

para sua regulação: definição de jurisdição, adequação do seu uso dentro da legislação atual de competição, responsabilidades, etc.

A opacidade é outro atributo que poderá derivar de complexidade ou de direitos proprietários, e que poderá estar presente em alguns sistemas de IA, com potencial de contribuição para dificultar a identificação de injustiças e, com isso, trazer barreiras para a identificação de que contrariam dispositivos legais, demandando, portanto, grande preocupação dos legisladores e dos reguladores. Conforme descrito em referencial sobre auditabilidade e em conformidade de soluções de IA no âmbito do Judiciário (ZAMRODAH, 2022), um sistema opaco é aquele no qual não se consegue explicar os detalhes ou as razões pelas quais a solução apresentou determinada decisão, e não outra. O dilema é que o uso de modelos de redes neurais profundas, aplicados em sistemas de IA, pode proporcionar excelentes resultados de acurácia sem que seja possível dar explicabilidade de como se chegou a eles, devido à rede complexa de camadas com cálculos matemáticos envolvendo a manipulação de inúmeras variáveis.

A literatura cita diversos casos identificados, nos quais decisões (BAROCAS *et al.*, 2019; MARTIN, 2021) com uso de técnicas de aprendizagem de máquina geraram prejuízos a determinado grupo, sendo comuns as citações de casos envolvendo o uso de reconhecimento facial em segurança pública, apoio à elaboração de sentenças judiciais, avaliação de riscos para concessão de crédito, seleção para entrada em universidades ou admissão em empregos, orientação a tratamentos de saúde, entre outros.

A Administração Pública brasileira vem, nos últimos anos, experimentando forte e crescente utilização de ferramentas digitais na gestão de políticas públicas de forma mais ampla, inclusive com a utilização de IA. Na esfera pública federal, há inúmeros relatos de uso de IA com aparente sucesso, bem como outros exemplos na esfera estadual (MCTI, 2021).

O relatório *Open Government Data Report*, de 2018, da OCDE (2018) destaca que a melhoria do acesso a dados governamentais poderá gerar novas formas de solucionar problemas da nossa sociedade, por meio da inovação e do

engajamento de partes interessadas do setor privado e público no processo de política de dados abertos. A premissa é a de que o uso dos dados abertos possa tornar-se o insumo necessário que proporcionará o desenvolvimento de aplicações e soluções úteis à sociedade, provocando benefícios econômicos de toda ordem, ao facilitar novas oportunidades de negócios e auxiliar a todos na tomada de melhores decisões com base em mais informações disponíveis.

A entrada em vigor da Lei de Acesso à Informação (LAI) (BRASIL, 2021) cujo objetivo é regulamentar preceitos constitucionais de maior transparência da Administração Pública, juntamente com a promulgação da Lei Geral de Proteção de Dados Pessoais (LGPD) (BRASIL, 2018) formam os contornos legais atuais de regulação para o setor público, ainda que de forma indireta, do uso de IA no país. Conjuntamente, trazem diversas obrigações e limites para o uso de informações pessoais e a transparência da atuação pública como regra. Nesse contexto de utilização de IA, destaca-se a obrigatoriedade de explicação clara e adequada a respeito dos critérios e dos procedimentos utilizados para uma decisão, se questionados por potencial interessado quando do uso de tratamento automatizado de dados pessoais para decisões que, porventura, possam ter afetado os interesses desse.

Outra questão fundamental, em especial no contexto de decisões que não sejam totalmente automatizadas, mas com uso de IA para seu suporte, é a possibilidade de redução de cautela ou pressuposição de maior certeza pelo decisor acerca de determinada informação quando essa é gerada com base em modelos quantitativos. Esse risco deriva da “expectativa” de que, por serem resultados gerados a partir de modelos matemáticos e estatísticos baseados em grande quantidade de dados representativos do universo de interesse, não se vislumbra que possam estar errados ou enviesados, seja por assimetria de conhecimento ou informações acerca de detalhes técnicos e limites inerentes à solução adotada, seja por ausência de informações complementares que auxilie a identificar o erro.

Simon Chesterman (2021) questiona que esse grande risco, proporcionado pela natural complacência humana frente a aceitar a sugestão de decisões geradas pela IA, diminuiu a capacidade humana de atuar corrigindo eventual

erro, mesmo quando o ser humano está inserido no processo decisório como revisor ou supervisor. Esse risco é tão relevante que o referencial de Singapura — *Model Artificial Intelligence Governance Framework Second Edition* — (SG:D., 2020) o traz como ponto de definição essencial de governança no uso de IA pela organização, no qual a determinação do nível apropriado de envolvimento humano na tomada de decisões com IA deve ser um processo iterativo, contínuo, alinhado aos valores organizacionais e documentado. O referencial *Guidance on IA and Data Protection* (ICO, 2020), elaborado no Reino Unido, ressalta, ainda, que, para um sistema não se caracterizar com decisão automatizada, de fato, é necessária uma intervenção humana significativa em cada decisão, sugerindo, inclusive, indicadores para esse monitoramento.

Consideradas as diversas especificidades relacionadas à adoção de um sistema de IA num processo decisório de gestão pública, que precisam ser consideradas dentro de seus contextos específicos, não é possível generalizar que ela sempre proporcionará melhorias para a gestão e para a sociedade, exigindo sempre um processo criterioso e consciente de governança na sua adoção frente às alternativas disponíveis.

Visando resumir e endereçar diretrizes de alto nível que têm sido incentivadas para os países e as organizações da Comunidade Europeia que desejam ampliar o uso de sistemas éticos de IA, cita-se um trecho do referencial de Orientações Éticas para uma IA de Confiança pela Comissão Europeia (2018):

Uma IA de confiança tem três componentes, que devem ser observadas ao longo de todo o ciclo de vida de um sistema: a) deve ser legal, cumprindo toda a legislação e regulamentação aplicáveis; b) deve ser Ética, garantindo a observância de princípios e valores éticos; c) deve ser sólida, tanto do ponto de vista técnico como do ponto de vista social, uma vez que, mesmo com boas intenções, os sistemas de IA podem causar danos não intencionais. Cada uma destas componentes é necessária, mas não suficiente, para alcançar uma IA de confiança. (EUROPEAN COMMISSION - GPAN AI, 2019).

Preocupações com a governança, com as bases de dados, com o monitoramento e com a avaliação desses sistemas, ao longo de todo o seu ciclo de vida, são eixos principais, com a inserção de ações e mecanismos específicos que se fazem necessários devido às particularidades e aos novos riscos que a nova tecnologia agrega ao processo de decisão.

Em que pese a discussão da existência de viés num processo decisório não ser um assunto novo, as preocupações atualmente aumentam muito em relevância e em potenciais consequências quando podem ser de alguma forma introduzidas por meio de tomadas de decisão automatizadas, parcialmente (*man-in-the-loop* e *man-on-the-loop*) ou totalmente (*man out of the loop*), por meio do uso de IA. Porém, o que justificaria essa ampliação de preocupações, já que o risco de vieses sempre esteve presente na construção de narrativas e justificativas que subsidiam uma decisão ou uma ação, como as relacionadas às políticas públicas? Visando endereçar soluções ao problema, o presente trabalho objetiva abordar, brevemente, o contexto de discussão atual sobre a regulação de IA no país que terá influência sobre as diretrizes de alto nível para o tratamento dos riscos de vieses no uso de IA pela Administração Pública.

REFERENCIAL TEÓRICO

No processo de tomada de decisão inerente à gestão de políticas públicas, os agentes políticos com o suporte das áreas técnicas trazem evidências acerca de um problema público que se deseja resolver, seu recorte de público-alvo e uma proposta de atuação estatal — desenho da política — pela qual se espera que, com a alocação de recursos para a implementação de ações, consiga-se, ao longo do tempo, eliminar ou mitigar causas ou consequências que justificaram a ação governamental. Porém, esse processo de criação, seleção e comunicação de evidências que justifiquem uma alocação de recursos públicos está longe de não estar sujeito a erros devido não somente à complexidade do contexto, por vezes envolvida, como também por direcionamentos propositais ou mesmo inconsciente das pessoas envolvidas nesse processo humano de compreender a realidade e emitir julgamentos e decisões.

PENSAMENTO CRÍTICO E MÉTODO CIENTÍFICO

As pesquisas desenvolvidas, por diversos pesquisadores com o uso de técnicas experimentais, para melhor compreensão de como o ser humano pensa e toma decisões, dentre os quais o nome mais conhecido é Daniel Kahneman (KAHNEMAN, 2011), clarifica as armadilhas e os grandes erros nos quais os humanos sujeitam-se nesses processos. O impacto dessas descobertas não somente rendeu prêmios Nobel de Economia a pesquisadores do tema, como ainda trouxe a grande relevância da necessária consciência das fragilidades do nosso processo cognitivo decisório frente às heurísticas e aos vieses, fato que reforça a grande relevância do tópico em questão na agregação de valor do uso de dados e da construção de evidências por pesquisadores para utilização pelos tomadores de decisão.

O grande avanço dos sistemas computacionais e de capacidade de processamento vem, nos últimos anos, proporcionando uma capacidade crescente de análise de grandes bases de dados para diversas finalidades. Há, inclusive, certa corrente que defende que vivemos a Era da Informação na qual os dados são o novo “petróleo”, porém, de forma similar, precisamos compreender que, para ele ter utilidade e valor, precisará ser adequadamente processado e refinado para determinados objetivos.

Se, por um lado, temos uma grande massa de dados criada e acumulada pela humanidade, por outro, a maior parte desses constituem-se de pouca valia e potencial fonte de erro para a compreensão dos fenômenos que buscamos compreender, mensurar e melhor decidir sob como agir sob esses.

Stephen Few (2019) defende que, para termos uma compreensão adequada dos dados quantitativos, nós precisamos pensar profundamente sobre eles, o que nos exige diversas habilidades, dentre as quais o conhecimento da área de domínio associado, relacionadas ao pensamento crítico e científico, da área de estatística, pensamento sistêmico, análise visual e pensamento ético. O autor cita uma definição trazida pelos educadores Richard Paul, Alec Fischer e Geral Nosich na qual o pensar criticamente deva ser um processo ativo, cuidadoso, persistente e focado na razão e nas evidências.

Richard Paul e Linda Elder (2014) defendem que um pensamento crítico deva ser composto de vários atributos presentes, por exemplo, a clareza, a acurácia, a precisão, a relevância, a profundidade, a lógica, a significância, a visão ampla e a justiça.

A aplicação do pensamento crítico, ao se analisar dados para se construir conclusões, opiniões e predições, busca, ao final, minimizar a exposição às fontes de erros (vieses e ruídos) que possam influenciar negativamente na decisão.

Stephen Few (2019) propõe, ainda, uma categorização e seleção de alguns vieses e suas origens que podem levar a erros no raciocínio quando saímos da simples descrição dos dados e utilizamos algum método para argumentarmos, analisarmos ou buscarmos prever algum comportamento futuro ou ainda explicar os dados com base no passado para uma tomada de decisão. Sua proposta agrupa vieses em erros por familiaridade, em erros estatísticos e em erros causais, que devem servir como pontos de atenção para os analistas e os cientistas de dados, muito embora o uso de algumas técnicas de IA adequadas possam mitigar alguns deles.

O método científico pode ser compreendido por um conjunto de princípios e práticas que direcionam o desenvolvimento da Ciência, podendo ser visualizados em três partes: as observações, que são as evidências na qual se baseia o conhecimento científico; a lógica que governa a interpretação das observações e os pressupostos, esses últimos similares a crenças, isto é, não podem ser provados, mas que se constituem em uma base fundamental de ponto de partida da ciência.

Exceto os pressupostos, tudo na ciência pode ser questionado e falseado a partir de novas evidências, possibilidade essa que muitos não compreendem bem ou entendem erroneamente como fragilidade do método científico, mas que, ao final, reside num grande pilar e na essência da teoria científica. Na visão teórica-filosófica de Karl Popper ([s.d.]), a busca em falsear seria um processo metodológico que visa recolher elementos que podem contestar a teoria, afirmando que uma teoria científica só deve merecer essa designação se for submetida a testes que possam contestá-la. Assim, quanto mais uma

teoria resistir às tendências do falseamento, mais ela será corroborável. De fato, considerando-se as diversas fragilidades cognitivas a que os humanos estão susceptíveis como fonte de erro, realizar testes que busquem refutar, e não propriamente confirmar a teoria proposta reduz que se incorra em vieses — o pensamento científico deve estar sempre procurando refutar a si mesmo.

Numa visão simples, o método científico seria entendido como a proposição de uma explicação para algo que não se compreende, por meio de construção de uma hipótese. A seguir procura-se realizar testes que buscam confrontar e questionar se a explicação proposta é válida frente às observações. A forma como a hipótese será testada dependerá da natureza do fenômeno e dos valores éticos que orientam a condução da pesquisa científica, podendo ser por meio de realização de experimentos ou por meio de estudos observacionais sem interferência do pesquisador.

Os pesquisadores formulam suas hipóteses, buscam, por meio de experimento ou observação, coletar dados, utilizam diversas técnicas e métodos sobre esses para, ao final, tentar refutar a hipótese — minimizando o risco do viés da confirmação.

DESAFIOS E OPORTUNIDADES DA INCORPORAÇÃO DE IA

A grande acumulação de bases de dados administrativos governamentais somada a outras que possam ser geradas por meio de experimentos ou estudos observacionais, ou até mesmo oriundas de compartilhamento com prestadores de serviços privados, geram grande oportunidade de produção de evidências para uso em políticas públicas. Porém, há uma quantidade numerosa de armadilhas e exigências de habilidades na produção de argumentos que auxiliem na compreensão dos fenômenos, isto é, evidências de qualidade. Com isso, as instituições que desejam contribuir com o processo de geração de evidências para uso de políticas públicas precisam fortalecer e sistematizar seus processos de geração de evidências, de forma que possam ser suficientes, confiáveis, fidedignas, relevantes e úteis à tomada de decisões em políticas públicas.

Dessa forma, o uso de IA pelo Estado tem o potencial de não somente aumentar sua eficiência com uso para processos internos, como fases de formulação de políticas públicas ou demais macroprocessos de suporte, como também para propiciar melhorias na prestação de serviços ao público-alvo das políticas públicas por meio de redução de erros nas decisões.

A literatura acerca do uso de sistemas de IA (DOMANSKI *et al.*, 2018) no processo de tomada de decisões em políticas públicas aponta diversos riscos e danos já identificados a grupos de cidadãos, por exemplo: impacto discriminatório (ZOOK *et al.*, 2017; MALEK, 2022), possibilidade de vazamentos de dados pessoais utilizados no ciclo de IA (ICO, 2020), desrespeito aos direitos fundamentais, justiça, entre outros.

Contudo, é necessário ter a clareza de que muitos deles já acontecem nas decisões de execução de políticas públicas, mesmo sem o auxílio de IA. A implementação de um criterioso sistema ético de IA pode ser a oportunidade de compreender vieses existentes não anteriormente visualizados pelo Estado e pela sociedade, bem como reduzir a variação inconsistente de decisões de diversos decisores sobre algo que deveria ter o mesmo tratamento. Essa última fonte de erro deve-se ao fato de os decisores humanos serem susceptíveis à influência de fatores e dados indevidos (ruídos) (KAHNEMAN *et al.*, 2021).

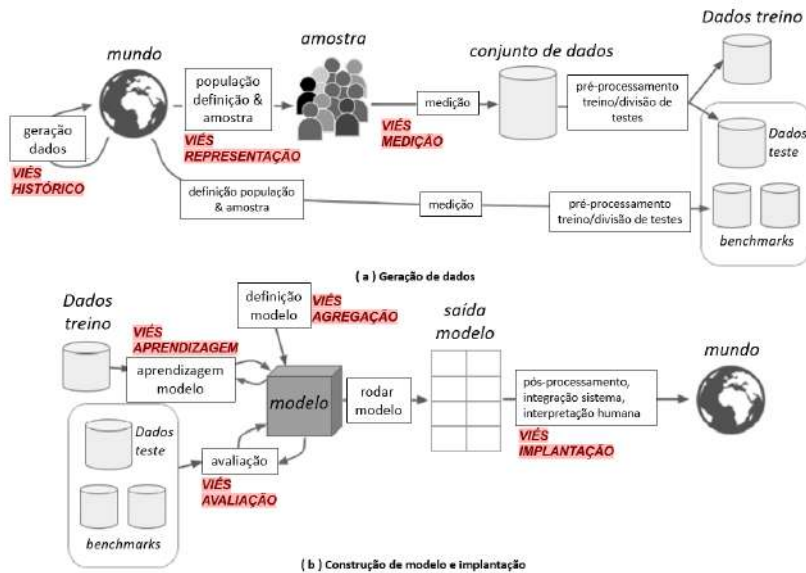
Acrescente-se, ainda, que a utilização de sistemas de IA poderá propiciar um salto de melhoria na qualidade, na uniformidade e nos custos das políticas públicas. Isso é possível por serem concebidos em algoritmos que estabelecem sequências definidas de ações e no uso de modelos desenvolvidos com o uso de métodos estatísticos e matemáticos e, com isso, menos sujeitos à dependência e à variabilidade da capacidade humana de decisão (vieses e ruídos) de diferentes pessoas inseridas numa cadeia do processo de decisão.

Porém, para que esses ganhos sejam assegurados sem a contrapartida de ocorrência de danos a terceiros, há de que se garantir um processo adequado de governança e princípios éticos guiando sua utilização na política pública. O processo de gerenciamento de riscos deve considerar não somente riscos cognitivos aos quais uma decisão humana estaria sujeita, como também novos

riscos específicos e inerentes ao de adoção de uma solução de IA para o apoio ou a tomada de decisão em determinado contexto específico.

Dentre as diversas formas de explicitar e agrupar vieses citados na literatura em sistemas de IA (MEHRABI *et al.*, 2021; Ntoutsi *et al.*, 2020), cita-se a sugerida por Suresh e Guttag (2021) pelos estágios do ciclo de vida de IA, que identificam 7 (sete) diferentes potenciais origens para viés cognitivo em um sistema de IA, conforme apresentado na Figura 1.

Figura 1 | Fases com Potencial de Geração de Viés para o Sistema De IA



Fonte: Tradução livre pelo autor com base no artigo citado.

Narayanan (2018) propôs uma consolidação de 21 diferentes formas de avaliar com critérios matemáticos consequências negativas de justiça a determinados grupos em um sistema de IA (*fairness*). Essa diversidade, bem como a impossibilidade de atendimento a todas as definições simultaneamente, corrobora que a discussão guarda não somente complexidade, como a necessidade de se fazer escolhas de projeto considerada a realidade

da aplicação e a impossibilidade de se garantir que um sistema opera sem nenhum viés. Ciente dos grandes desafios e das oportunidades, grande parte dos países e das instituições públicas e privadas tem incentivado uma visão principiológica, um modelo de regulação responsiva e um forte processo de conscientização dos atores envolvidos com toda a cadeia de produção e utilização dos sistemas de IA.

O referencial *Model Artificial Intelligence Governance Framework — 2ª edition* (SG:D.,2020) traz uma compilação dos princípios que se esperam estar associados ao desenvolvimento de uma IA ética: Responsabilidade, Responsividade, Acurácia, Auditabilidade, Explicabilidade, Alinhado aos direitos humanos, Inclusividade, Justiça(*fairness*), Progressividade, Robustez, Segurança, Sustentabilidade, Governança e Transparência.

Visando que os sistemas atendam aos diversos princípios, vários estudos, avaliações e decisões internas de projeto devem ser realizados a depender do contexto e, por isso, é fundamental a documentação de todo o processo. A adequada documentação (GAO, 2021) perpassa desde a definição clara dos objetivos do uso de IA, políticas e níveis de autoridade aprovadora aplicáveis, critérios de eficácia mínima e limites de desvios toleráveis ao longo do ciclo de uso, gerenciamento de riscos, requisitos de explicabilidade, avaliação do impacto do uso de dados pessoais, justificativa de certas escolhas em *trade-offs* durante o projeto, processo de monitoramento e de avaliação, inclusive com os resultados dos já realizados para posterior comparação. Busca-se, com isso, propiciar auditabilidade, responsividade, explicabilidade e, ao final, uma melhor tomada de decisão pública de forma transparente, legal, confiável e centrada no ser humano.

Em geral, a maioria dos países tem buscado uma regulação responsiva com o foco no estabelecimento de incentivos à prevenção e à conformidade regulatória, de forma que não se crie um freio ou um desincentivo aos avanços no uso dos sistemas de IA nos diversos setores. Dessa forma, tem-se buscado traçar princípios e limites norteadores de forma participativa entre governo e mercado que deverão ser respeitados e perseguidos pelos responsáveis pelo desenvolvimento e pelo uso de sistemas de IA. Além disso, tem-se estabelecido,

de acordo com o setor, estruturas de governança claras que tenham *expertise* sobre o tema de forma que possam atuar diligentemente e independentemente nos casos de necessidade de verificação futura de eventual não atendimento das diretrizes e das orientações.

Quanto ao melhor modelo de estrutura de governança e regulação a ser criada em cada país, ou mesmo no incentivo de criação de padrões internacionais por meio de agências multilaterais (CHESTERMAN, 2021), todos são pontos que ainda estão em discussão. Alguns países vêm discutindo ou adotando a atribuição de novas funções e papéis a instituições já existentes em vez de criar novas estruturas para lidar com IA, até por entenderem que IA poderá ter aplicabilidade transversal, perpassando por quase todas as áreas, como finanças, saúde, jurídica, ambiental, entre outras, o que acaba por exigir um detalhado conhecimento específico de cada setor para uma abordagem de regulação apropriada. Outro ponto importante é que, no uso de IA, é muito comum que haja uso de dados pessoais, tema sobre o qual alguns países já possuem regulação específica.

Em geral, os tomadores de decisão com o suporte de IA, denominados de revisores, não são as mesmas pessoas que processam os dados, analisam-nos e selecionam técnicas adequadas para os objetivos pretendidos. De forma que o revisor saiba da falibilidade do sistema de IA e compreenda quando é mais provável que ele venha a cometer erro em sua sugestão, permitindo ao revisor decidir acertadamente, faz-se necessária uma adequada redução de assimetria entre quem desenvolve e quem utiliza a IA para a decisão. Com isso, a forma e o processo de comunicação assumem papéis fundamentais para mitigar erros e eventuais danos a terceiros, devendo essa relevância ser reconhecida e os riscos mitigados de forma consciente pela organização que adote o uso de IA em decisões.

ANÁLISE

Busca-se, no presente trabalho, realizar um olhar comparativo do estágio nacional com o identificado nos Estados Unidos e no Reino Unido acerca

da discussão regulatória da temática do uso de IA. Adotou-se um escopo mais restrito e limitado aos eixos temáticos horizontais previstos na EBIA de “legislação, regulação e uso ético” e “governança” de Inteligência Artificial, no recorte de eixo vertical “aplicação no poder público”, tendo em vista o problema definido.

A revisão de literatura não somente aponta alternativas viáveis para melhor utilização dessas tecnologias em equilíbrio com a preservação da justiça nos seus resultados, como também externaliza o intenso movimento de diversos países e organismos multilaterais para exercerem seu papel de liderança nas discussões, envolvendo todas as partes interessadas — setor privado, academia, setor público e entidades representativas da sociedade civil.

Uma análise rápida das propostas de iniciativas apresentadas pelo Brasil, pelo Reino Unido e pelos Estados Unidos ao levantamento feito pela OCDE deixa claro que há diferenças no estágio de maturidade das discussões entre os países selecionados, não somente pela quantidade, mas pela abrangência e pela profundidade das propostas informadas.

Acerca de iniciativas para Orientação e Regulação, o Brasil informou ter 2 (duas) iniciativas, enquanto o Reino Unido e os Estados Unidos já contam com 22 (vinte e duas) e 18 (dezoito) respectivamente. Em relação ao tema da Governança, o Brasil conta com 4 (quatro) iniciativas, enquanto o Reino Unido já possui 20 (vinte) e os Estados Unidos, 18 (dezoito) diferentes iniciativas.

Verifica-se que, nos EUA e no Reino Unido, já existem iniciativas focadas em alguns setores, como o mercado nascente de veículos autônomos, as finanças e as seguranças pública e militar, enquanto no Brasil as iniciativas de legislações e estratégias são mais gerais, algumas delas indiretamente relacionadas à IA, mas não específicas, como as voltadas à transformação digital, à internet das coisas e à proteção de dados pessoais.

Considerada a grande variedade de sistemas de IA, bem como a dificuldade de buscar regular simplesmente tecnologias desconectadas de suas aplicações concretas, os governos dos países mais avançados no tema vêm buscando conjuntamente com o mercado privado e com a sociedade criar e

disseminar discussões, conhecimento, boas práticas e princípios que os utilizadores dessas tecnologias deverão compreender, assimilar, dentro de suas culturas organizacionais e garantir que estão em conformidade para minimizar eventuais danos a terceiros, alinhado com uma visão de regulação responsiva.

REALIDADE BRASILEIRA

A velocidade de adoção do que vem sendo denominado como a Quarta Revolução Industrial (MCTI, 2021) e que está promovendo mudanças significativas na tecnologia da informação inteligente (TI inteligente) ganhará mais impulso com a implementação da conexão 5G no país e com a expansão da infraestrutura de conectividade para suportar a adoção de equipamentos dotados de sensores no conceito de Internet das Coisas — IoT.

Ao final de 2019, o Ministério da Ciência Tecnologia e Inovações e Comunicações (MCTIC), à época, dentro de suas competências institucionais, lançou consulta pública para subsidiar a elaboração da EBIA, que foi positivada pela Portaria MCTI n. 4.617, de 6 de abril de 2021 (MCTI, 2021). Após diversas fases, em abril de 2021, o Ministério da Ciência Tecnologia e Inovações (MCTI) instituiu a EBIA, composta de 9 (nove) eixos temáticos, 3 (três) transversais e 6 (seis) verticais.

A EBIA, dentro da discussão acerca de governança de IA, trata da questão da relevância da análise de riscos associada às legislações modernas e mais especificamente da elaboração dos relatórios de impacto de proteção de dados (RIPD) para avaliar questões de justiça, direitos humanos ou de outras considerações para implementação de IA, como sendo instrumento de *accountability* e documentação para auditabilidade, explicabilidade e transparência. Cita, ainda, que cabe à Autoridade Nacional de Proteção de Dados (ANPD) a responsabilidade para editar diretrizes para orientar a elaboração desse documento especialmente relevante para incentivar os controles preventivos e detectivos planejados e executados, além dos critérios de decisão adotados com suas justificativas em relação à proteção de dados pessoais.

Posteriormente, o MCTI estabeleceu uma estrutura de governança da EBIA com a publicação de um regimento interno, a ser exercida por um Comitê de Governança composto por: MCTI; Rede MCTI/EMBRAPII de Tecnologias e Inovação Digital e instituições convidadas. A estrutura de governança criada aprova então a criação de um Subcomitê Temático para cada eixo da EBIA, com a escolha de coordenadores titulares e suplentes, dentre representantes da esfera pública e privada, visando dar desdobramentos à política.

O Tribunal de Contas da União, considerando a grande relevância da temática sob diversas perspectivas e pela aplicação em diversas instâncias governamentais, realizou auditoria de levantamento do uso de IA nas organizações da Administração Pública Federal (APF) (BRASIL, 2022b), culminando no Acórdão n. 1.139/2022.

Destacam-se em relação ao acórdão, os diversos achados de auditoria relativos à política pública da EBIA na qual fica claro a ausência de realização de uma análise *ex-ante* pelos seus gestores responsáveis, resultando em uma política pública sem seus elementos, instrumentos e características fundamentais e esperados que possibilitem adequado monitoramento e avaliação ao longo de sua implementação.

Recentemente, o Senado Federal instituiu uma Comissão Temporária Interna destinada a subsidiar a elaboração de minuta de substitutivo para instruir a apreciação de projetos de lei sobre IA que têm como objetivos estabelecer princípios, regras, diretrizes e fundamentos para regular o uso da nova tecnologia no Brasil (BRASIL, 2022a). O cronograma de seus trabalhos previa que uma proposta fosse apresentada em agosto, após a realização de diversas audiências públicas e seminário internacional com especialistas, porém esse sofreu prorrogação para dezembro do presente ano (2022).

No país, a proposição da EBIA guarda forte conexão com as diretrizes da OCDE, de forma coerente ao fato de o país ter sido signatário em 2019 dos Princípios da OCDE sobre Inteligência Artificial, o que, a priori, trazia boas expectativas iniciais acerca de seus resultados, porém que somente serão concretizados se as diversas ações propostas dentro dos eixos lograrem êxito

ao longo do tempo. De certa forma, os achados do trabalho de auditoria do TCU enfraquecem essas expectativas no curto prazo pela fragilidade de governança e gestão da política proposta.

O contexto de diversas iniciativas de uso de sistemas de IA no âmbito do Poder Judiciário, somado à lacuna normativa sobre o tema, levou o Conselho Nacional de Justiça (CNJ), dentro do seu papel de supervisão no Poder Judiciário, a editar a Resolução n. 332/20 (CNJ, 2020) para orientar boas práticas e estabelecer diretrizes, alçadas acerca de riscos éticos, transparência e governança no uso de Inteligência Artificial no Poder Judiciário.

Considerando princípios trazidos em diversos referenciais de IA, bem como as características de técnicas de aprendizagem de máquina demandante de grandes bases de dados e, especificamente para uso governamental, de base de dados pessoais, não há como se avaliar o contexto legal, regulatório e de governança no setor público de forma integral sem verificar também diversos subsistemas associados — o de transparência pública e o de proteção de dados pessoais.

É inegável o avanço legal e mesmo instrumental no Brasil na implementação do princípio da transparência pública, desde a promulgação da LAI, com papéis, responsabilidades e processos constituídos e sedimentados, em especial, na esfera federal.

Ao longo dos anos, dados como remuneração dos servidores públicos federais, planejamento e execução orçamentária e financeira do governo federal, relatórios fiscais que mostram as trajetórias dos diversos tipos de dívidas, gastos específicos de políticas públicas, incluindo visões regionalizadas com a identificação de objetos, dentre diversas outras informações, passaram a ser disponibilizadas por meio de portais no atendimento da transparência ativa. Soma-se a esse esforço crescente ao longo dos anos o aprimoramento do processo de transparência ativa por meio da implementação da política de dados abertos pelos órgãos públicos, importante caminho para o fortalecimento da disponibilização e da utilização de dados públicos no controle social. Nos casos nos quais a informação não se encontra publicizada, há ainda

o caminho da transparência passiva por meio do atendimento aos pedidos feitos por meio da LAI.

Quando o foco de análise é a implementação do princípio de proteção de dados pessoais, precisa-se registrar que, muito embora houvesse previsões limitadas acerca do tema em outras legislações, como a própria Constituição Federal de 1988, a Lei n. 10.406/2002 — Código Civil—, dentre outras, a LGPD é o principal marco regulatório. A LGPD, inspirada na legislação de proteção de dados europeia *General Data Protection Regulation* (GDPR), possui, mesmo após 4 (quatro) anos de sua promulgação, muitos desafios relevantes a serem superados para sua implementação pelos agentes regulados, conforme identificado por auditoria (BRASIL, 2022c), recentemente concluída pelo TCU e constante do Acórdão n. 1.384/2022.

A auditoria apontou uma situação de alto risco à privacidade dos cidadãos que têm dados pessoais coletados e tratados pela Administração Pública Federal. O resultado do trabalho comparou as organizações auditadas quanto ao nível de adequação à LGPD, chegando à conclusão de que 17,8% estão no nível inexpressivo; 58,9% estão no nível inicial; 20,4% estão no nível intermediário e 2,9% estão no nível aprimorado. Esses números suportaram a opinião de que se tem uma situação de alto risco à privacidade dos cidadãos que possuem dados pessoais coletados e tratados pela Administração Pública Federal.

Destaca-se, dentre as recomendações, a relacionada à temática de governança sobre o tema, que foi para que sejam adotadas “as medidas necessárias para alterar a natureza jurídica e promover a reestruturação organizacional da Autoridade Supervisora de Proteção de Dados — DPA (*Data Protection Authorities*) nacional, conferindo-a o grau de independência e os meios necessários para o pleno exercício de suas atribuições”.

Os princípios trazidos e definidos pela LGPD, pela proximidade que guardam com a temática de IA, são: finalidade, adequação, necessidade, livre acesso, qualidade dos dados, transparência, segurança, prevenção, não discriminação, responsabilização e prestação de contas.

Um olhar cuidadoso sob os princípios contemplados permite compreendê-los como entrelaçados para propiciar o atendimento de princípios éticos de um sistema de IA, em especial, dos que fazem uso de aprendizagem de máquina com uso de dados pessoais.

Por exemplo, a garantia que os titulares dos dados tenham o livre acesso para saber quais dados pessoais estão sendo tratados e a transparência com informações claras, precisas e facilmente acessíveis sobre a realização do tratamento contribuem para garantir que o sistema de IA atenda aos princípios éticos da transparência, da justiça e da acurácia. Por decorrência da compreensão dos efeitos sob determinado direito que uma decisão automatizada com uso de IA esteja ocasionando, o titular que se sinta prejudicado pode solicitar correção/atualização de dados equivocados ou desatualizados, contribuindo também com o princípio da qualidade dos dados previsto pela LGPD em última instância.

Outra questão menos tangível é que a existência da obrigação dessas garantias disciplina o controlador (responsáveis pelas decisões referentes ao tratamento de dados pessoais) acerca da necessária consciência dos potenciais efeitos sobre os direitos dos cidadãos, deixando claro a sua responsabilidade em avaliar, mitigar e documentar os riscos existentes em todas as fases do ciclo, reforçando a visão de responsabilidade, responsividade, explicabilidade e auditabilidade, também princípios de IA ética e confiável.

Considerada a sua importância, buscou-se analisar as orientações/regulamentações do RIPD — Relatórios de Avaliação de Impacto de uso de dados pessoais. Na LGPD, sua definição permite inferir a intenção do legislador de que os impactos do tratamento do dado pessoal estejam associados a sua finalidade de uso. Isso traz a necessidade de que impactos acerca de potenciais vieses que possam causar prejuízos a alguns grupos na preservação de seus direitos em sistemas de IA sejam avaliados no RIPD e, por isso, as orientações dadas em eventos e documentos devam deixar isso de forma clara, visando incentivar uma atuação dos controladores e dos operadores mais abrangente e consciente a riscos mais específicos e associados à IA.

A LGPD atribui à ANPD o papel para editar regulamentos e procedimentos sobre esse instrumento de grande valia para o processo de adoção de IA para os casos em que o tratamento representar alto risco à garantia dos princípios gerais de proteção de dados pessoais. Embora constasse da agenda regulatória da agência prevista para 2021–22, a regulamentação do instrumento aparece como em estágio de “andamento” na proposta de agenda regulatória para consulta pública para o biênio de 2023–24. Ressalta-se também o interesse da agência de dedicar-se ao tema de Inteligência Artificial e suas implicações constante de sua agenda para o próximo biênio, sendo uma oportunidade para melhor orientar acerca dos objetivos e dos aspectos relevantes da elaboração do RIPD dentro desse contexto tecnológico .

Em agosto de 2020, o Comitê Central de Governança de Dados aprovou, por meio da Resolução CCGD n. 4, de 14 de abril de 2020, um guia de boas práticas da LGPD (Comitê Central de Governança de Dados, 2020), porém sem adentrar explicitamente em especificidades relacionadas à produção do RIPD quando do uso de dados pessoais em sistemas de IA.

Atualmente, há um modelo de RIPD (SGD/ME, 2020) disponibilizado pela Secretaria de Governo Digital (SGD/ME) com orientação dos campos e dos tipos de informações a serem disponibilizadas, muito alinhado ao que se adota em outros países sujeitos à GDPR num olhar rápido. Considerada a abordagem que vem sendo adotada de regulação responsiva, orientações claras e modelo tomam grande relevância para o guiamento dos atores no processo de adoção das boas práticas que minimizam riscos e, por isso, ele foi objeto de análise. Adotou-se, para fins de comparação dos campos, o modelo sugerido pelo *Information Commissioner’s Office* (ICO), no âmbito do Reino Unido, e as orientações acerca do instrumento disponibilizadas no seu sítio eletrônico com o modelo proposto pela SDG.

Os campos e as orientações gerais de seu preenchimento, constantes dos documentos, não trazem nenhuma diferença entre eles numa rápida comparação. Porém, quando se traz ao contexto de análise alguns delineadores que devem guiar uma regulação responsiva, observa-se alguns pontos como oportunidade de melhoria no modelo nacional para adequada persuasão

junto aos atores. No campo denominado “Identificar e avaliar os riscos” no modelo nacional, são sugeridos em rol exemplificativo de catorze riscos, não havendo nenhum mais relacionado às discussões de vieses em IA ou em justiça. Essa escolha para o modelo, ainda que exemplificativa, pode deixar frágil a necessidade de induzir a consciência para identificar e tratar os riscos de vieses de IA, para uma consciente adoção dessa tecnologia no processo de tomada de decisões. A maturidade dos órgãos governamentais, bem como do setor privado, na adoção de sistemas éticos de IA ainda é incipiente em relação à consciência necessária para o tratamento da temática e, por isso, seria de grande relevância que o modelo pudesse contemplar explicitamente alguns riscos relacionados a vieses, porque esses têm o potencial de afetar negativamente direitos fundamentais no tratamento dos dados pessoais aplicado à finalidade do uso do sistema de IA.

Um outro ponto que se identificou como importante oportunidade de melhoria no modelo do instrumento, considerando os princípios de auditabilidade, explicabilidade e responsividade que devem guiar o desenvolvimento de um sistema ético de IA, é a forma em que se registra o rito dos aprovadores exigidos no processo. No documento estrangeiro, ficam claramente identificados os responsáveis, bem como os respectivos posicionamentos técnicos acerca dos potenciais impactos decorrentes das decisões tomadas na proposição do uso de IA para o processamento de dados pessoais para uma determinada finalidade. Discordâncias relevantes que possam surgir no processo altamente complexo e técnico, com eventuais justificativas na tomada de decisão para seguir em frente, como a não adoção de sugestão de uma medida de mitigação de risco feita pelo *Data Protection Officer (DPO)*, ficam documentadas e explicitadas. Diferentemente, no modelo nacional sugerem-se apenas as assinaturas, formato este que pode desincentivar a responsabilidade e o *accountability*, aumentando o apetite ao risco nas decisões. Considerando que o documento deva ser um instrumento vivo e atualizado com as mudanças implementadas durante todo o ciclo da IA, essa rastreabilidade do exercício dos diversos papéis é fundamental não somente para registrar e demonstrar uma atuação responsável e diligente, como também para contribuir fortemente

para a adequada documentação do processo de desenvolvimento do sistema de IA e para o aprendizado organizacional.

O arcabouço legal pátrio existente atribui à ANPD, dentre as diversas competências relacionadas à proteção de dados pessoais, a competência para atuar como instância com papel de auditoria nos setores público e privado quando decorrente do não atendimento pelo controlador de prestação de esclarecimentos de “informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada” ao titular de dados que se sentiu prejudicado por decisão. Dentro do desenho de regulação responsiva, essa competência tem papel dissuasório relevante para persuadir os regulados a adotarem as boas práticas importantes para uma adequada adoção de sistemas éticos de IA no país. Porém, para o efetivo exercício dessa competência, a ANPD deverá dispor de competências coletivas adequadas e não necessariamente disponíveis dentro de seu quadro de recursos humanos, restando como alternativa a articulação de acordos de cooperação com demais atores públicos que detenham essa expertise num curto prazo.

Decorridos 4 (quatro) anos da promulgação da LGPD e de contínuo crescimento de uso de sistemas de IA, o país ainda carece de melhores definições acerca de papéis e de liderança efetiva para regular tal e promissora tecnologia pelas organizações. Considerada a dimensão que o endereçamento do tema exige, a abrangência e a multidisciplinariedade de conhecimentos, bem como ainda o nível de aprofundamento técnico necessário, estabelecer um adequado desenho de governança pública não será uma decisão trivial no Brasil e ainda se encontra em muitos países pendente de uma definição mais clara.

O MCTI busca exercer papel de liderança nas discussões e na interlocução com as diversas partes relacionadas e afetadas acerca de como essas tecnologias deverão ser abordadas na EBIA, porém o desenho de política pública, conforme já apontado, ainda carece de uma definição mais objetiva da situação atual e a que desejamos num futuro definido. A 5ª Reunião Extraordinária do Comitê de Governança da EBIA, realizada em fevereiro de 2022, trouxe como pauta o processo de revisão da Estratégia Digital — E-Digital, que se encontra no momento em revisão. É importante pontuar que a Casa Civil exerce a lide-

rança e o MCTI, a Secretaria Executiva, na governança do Sistema Nacional para a Transformação Digital, tendo a temática grande sobreposição e sinergia com a temática de Inteligência Artificial. É uma oportunidade relevante para que se busque otimizar e harmonizar a alocação dos recursos nas prioridades relacionadas para o enfrentamento do tema, no qual a IA pode ser enxergada como tecnologia relevante numa visão de uma sociedade digital.

Dessa forma, no momento atual, as utilizações de IA pela Administração Pública sob a perspectiva federal, excetuando o Poder Judiciário, padecem de maior clareza acerca da governança, modelo regulatório e princípios a serem perseguidos para os projetos de inserção de IA no processo decisório, ainda que o país tenha aderido aos Princípios da OCDE sobre Inteligência Artificial em 2019.

A governança e a regulação do tema no país, consideradas as iniciativas realizadas e outras em curso, em termos de maturidade, pode ser descrita ainda como incipiente frente aos EUA e ao Reino Unido, pelo menos sob uma ótica de desenho de papéis e ações, sem que se possa adentrar numa comparação de efetividade. Ressalta-se que, atualmente, apenas o Poder Judiciário dispõe de um processo de governança e diretrizes estabelecidos por meio de resolução do CNJ (2020) para o uso de IA pelos órgãos que o compõe.

REALIDADE NOS ESTADOS UNIDOS DA AMÉRICA (EUA)

A posição pendular em favor dos Estados no desenho federativo norte americano reflete fortemente sob a forma de legislar, regular e, mesmo, decidir os litígios acerca de diversas temáticas.

As fortes influências de políticas liberais vêm ditando a forma de legislar e regular a IA de maneira bem cautelosa, considerado não somente o histórico do seu federalismo de cooperação — com menor centralização de competências na esfera federal —, mas também o fato de que as maiores empresas que desenvolvem a fronteira das tecnologias relacionadas à IA são norte-americanas.

Embora haja guias regulatórios em determinados setores emitidos por agências, o país ainda não possui uma lei ou outro normativo federal com o papel de regular de forma mais ampla o tema de IA.

No processo de fomento à discussão sobre o tema, citam-se como relevantes os atos normativos expedidos pelo Chefe do Poder Executivo, denominados de *Executive Order* (EO) 13.859 (ESTADOS UNIDOS DA AMÉRICA, 2019) e 13.960 (ESTADOS UNIDOS DA AMÉRICA, 2020). Em visão agregada, eles lançaram a estratégia do Governo Federal para IA — *American AI Initiative* —, a ser coordenada pelo *National Science and Technology Council* (NSTC), determinaram ações voltadas a mapear todas as iniciativas em curso de IA dentro das agências, demandaram que o Departamento de Comércio americano (*U.S. Department of Commerce — DoC*), por meio do *National Institute of Standards and Technology* (NISTP), apresentasse um plano para a atuação federal no desenvolvimento de padrões técnicos e ferramentas de apoio a sistemas confiáveis, robustos e de confiança, padronizaram e definiram alguns papéis para a Administração Pública Federal no uso de IA e estabeleceram que os princípios da legalidade e do respeito aos valores nacionais, do direcionamento pelo propósito e desempenho, da precisão, da confiabilidade e da eficácia, da segurança e da resiliência, da compreensibilidade, da responsabilidade e da rastreabilidade, do monitoramento regular, da transparência e da prestação de contas devam ser perseguidos para o uso de IA dentro da Administração Pública (*Principles for the use of AI in Government*). Demandaram, ainda, que o *Office of Management and Budget* (OMB) publicasse orientações às agências para melhor apoiar o uso de IA, bem como incentivar a adoção de padrões voluntários e consensuais desenvolvidos com a participação da indústria.

Posteriormente, em novembro de 2020, o OMB emitiu o memorando M-21-06 às agências com o propósito de guiá-las para como atuarem na regulação do uso de IA. A orientação adota tom demasiadamente cauteloso e excessivamente focado em argumentar que a regulamentação não deva impedir a inovação e a implantação, fato que lhe rendeu críticas (ENGLER, 2022), por não sobrepesar de forma equilibrada os ganhos enaltecidos frente aos eventuais danos que o uso da tecnologia possa causar. O documento

clarifica a decisão de que a regulação, quando necessária, deva ser específica para os setores, sobrepesando o tipo de aplicação da IA, em vez de políticas mais abrangentes que possam não fazer sentido em todo o amplo espectro do uso da IA. Sugere também a priorização de proteções mais fortes para IA que demonstrem potencial de maior risco, bem como indica os princípios e as práticas regulatórias que entende relevantes a serem perseguidos pelas agências na regulação: confiança pública na IA, participação pública, integridade científica e qualidade da informação, avaliação e gerenciamento de riscos, custos e benefícios, flexibilidade, justiça e não discriminação, divulgação e transparência, segurança e proteção e coordenação interagência. Sugere também atuação não regulatória nos casos em que entender que a regulação já existente é suficiente frente aos novos riscos trazidos pela incorporação de IA, bem como a adoção de: política de desenvolvimento de guias e referenciais específicos para o setor, criação de experimentos e programas pilotos com a permissão para a realização de testes sujeitos a requisitos regulatórios específicos para melhor compreensão do uso de IA (*sandbox* regulatório), incentivo à elaboração de normas e referenciais consensuais voluntárias.

O NIST apresentou um plano (NIST, 2019) no qual identificou 9 (nove) áreas de foco para serem priorizadas pelas normas voltadas para IA: conceitos e terminologia, dados e conhecimento, interações humanas, métricas, trabalho em rede, teste de desempenho e metodologia de relatórios, segurança, gerenciamento de risco e confiabilidade. Sugere que os padrões de confiabilidade devam incluir orientação e requisitos de precisão, capacidade de explicação, resiliência, segurança, confiabilidade, objetividade e segurança, bem como estejam alinhados com políticas e princípios do governo dos EUA, que incluem preocupações acerca de questões sociais e éticas, governança e privacidade.

O NIST também está elaborando, desde 2021, um Referencial de Gerenciamento de Riscos de IA — *AI Risk Management Framework* —, com previsão de publicação para 2023, com duas versões preliminares já disponibilizadas para discussão e aprimoramentos. A sua segunda versão para revisão (NIST, 2022) está estruturada em 4 (quatro) funções: Governar, Mapear, Medir e Administrar, que se desdobram em categorias e subcategorias. Consta ainda, como pontos

interessantes propostas de tarefas e responsabilidades a diferentes atores dentro de um ciclo de IA, bem como a identificação de riscos diferenciados de IA dos riscos tradicionais de sistemas de tecnologia da informação.

Em 2021, por meio do *National Defense Authorization Act for Fiscal Year 2021*, criou-se o *The National Artificial Intelligence Initiative Office (NAIIO)*, dentro da estrutura do *Office of Science and Technology Policy (OSTP)*, na Casa Branca para supervisionar e coordenar as estratégias de IA em âmbito federal.

Em atenção aos impactos que o uso da tecnologia de IA trará não somente para a gestão, mas também os novos desafios para os processos de auditoria, o *U.S. Government Accountability Office (GAO)* — estrutura similar ao Tribunal de Contas da União no país — publicou, em 2021, um referencial para uso das agências federais e outras entidades sobre o tema. O *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities* (GAO, 2021) buscou identificar práticas-chaves para promover o uso responsável da IA por agências federais e outras entidades envolvidas no projeto, no desenvolvimento, na implantação e no monitoramento contínuo dos sistemas de IA. O referencial é organizado com foco em torno de 4 (quatro) princípios complementares: governança; dados; desempenho e monitoramento, sendo estruturado por meio de questionamentos a serem feitos aos diversos atores dentro do seu papel no ciclo de IA.

Pesquisadores das Universidades de *Stanford* e de *New York* elaboraram o documento *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* (ENGSTROM *et al.*, 2020) no qual trazem o resultado da análise de 7 (sete) estudos de casos relacionados ao uso de IA para a execução civil, execução híbrida civil/criminal, adjudicação formal, adjudicação informal, análise regulatória, engajamento público e prestação de serviços públicos nos EUA e um compilado dos usos de IA no governo.

Dessa forma, embora os EUA possuam grande expertise e conhecimentos técnicos acerca de IA, atualmente não há uma regulação federal (SUSSMAN; MCKENNEY; WOLFINGTON, 2022) para uso pela Administração nem pelos agentes regulados. Contudo, em alguns setores específicos, há medidas não

regulatórias implementadas por algumas agências com o objetivo de reduzir riscos. Importante trazer ao contexto que, dentre outras propostas, tramita no Congresso, desde 2019, reapresentada em 2022, a proposta de lei *Algorítmica Accountability Act of 2022* para positivar uma diretriz transversal federal.

Sob a ótica mais delimitada de uso de IA pelas organizações públicas, existe a orientação de que esses sistemas sigam os princípios elencados na EO 13.960. Outro ponto positivo que merece ser destacado é o fato de o órgão responsável pelas auditorias externas (GAO) emitir um referencial com boas práticas, já sinalizando o que poderá ser demandado na análise de cada caso concreto em uma futura auditoria, porém a grande fragmentação de agentes reguladores e a não emissão de norma transversal, principiológica e flexível induzindo de forma mais clara uma regulação responsiva delimitam o contexto no qual uma necessária atuação do regulador junto ao agente regulado na apuração de responsabilidade possa ser questionada no âmbito administrativo, bem como no âmbito judicial pela fragilidade regulatória.

Num contexto mais amplo, relacionado à IA e seus impactos, é importante saber que nos EUA também não existe uma legislação federal de proteção de dados pessoais voltada para regular a transparência pública, nem mesmo um único órgão responsável pela condução dessas políticas.

Importante ainda destacar que, diferentemente de outros países, nos EUA o direito à privacidade é visto como instrumento de proteção dos consumidores, e não como direito fundamental, fato essencial que influencia o papel do regulador e resulta em práticas e em diferenciação entre as DPAs e as outras reguladoras que abordam o tema de privacidade no país. Coexistem ainda centenas de outras legislações federais e estaduais com essa finalidade dentro de aplicabilidades específicas.

A análise retrospectiva de atos e ações voltados para regular IA nos EUA deixa evidente que a governança e o avanço normativo da tecnologia seguem dentro de uma cultura regulatória já, utilizada para outras temáticas, de atuação descentralizada e buscando, sempre que possível, uma quase autorregulação pelos próprios setores. Fruto do alto conhecimento tecnológico sobre o tema,

há bons referenciais sendo desenvolvidos para uso setorial ou mesmo mais amplo por instituições normatizadoras.

Esses instrumentos, ainda que não cogentes para as partes, poderão servir como insumo para disseminação de conhecimento e boas práticas, dentre as quais o incentivo à adoção do gerenciamento de riscos de IA, além dos já tradicionais riscos corporativos da organização.

REALIDADE NO REINO UNIDO (UK)

O UK vem construindo a agenda regulatória sobre o uso de IA de forma muito ativa, tendo sido uma das justificativas para sua escolha no uso comparativo no presente trabalho. Muito embora ele tenha saído oficialmente da União Europeia, por meio do conhecido processo *Brexit*, ao final de janeiro de 2020, sua permanência no bloco durante vários anos bem como os fortes laços comerciais e de negócios influenciaram fortemente o processo de regulação no país.

Enquanto fazia parte da União Europeia (UE), a regulação transversal em relação à proteção de dados privados era a GDPR, porém, com o seu processo de saída, o parlamento aprovou o *Data Protection Act 2018* (DPA, 2018). A legislação aprovada possui algumas diferenças, embora seja bem alinhada à legislação anterior do bloco, estabelecendo de forma similar princípios a serem seguidos por quem processa informações relacionadas a indivíduos, direitos aos proprietários dos dados, aplicáveis tanto ao setor privado quanto pelo uso governamental bem como obrigações às partes. O órgão ao qual se atribui a orientação, a supervisão e a fiscalização pela implementação do DPA 2018 no setor público e privado é o *Information Commissioner's Office (ICO)*, órgão independente dentro da estrutura do *Department for Digital, Culture, Media and Sport (DCMS)*.

O ICO, isoladamente, ou por meio de parceria com o *Alan Turing Institute (The Turing)* — instituto voltado para ciência de dados e IA — tem promovido a publicação de muitos guias e orientações, fruto de pesquisas envolvendo os

diversos setores, voltados para a compreensão dos riscos para se implantar sistemas de IA ética. Destacam-se os seguintes materiais disponibilizados pela qualidade das informações: *Project ExplAIIn Report*, *Guidance on AI and Data Protection*, *AI and Data Protection Risk Toolkit*, *Guide to data protection*, *Accountability and governance Data Protection Impact Assessments (DPIAs)*.

As diretivas da GDPR exigem a elaboração do DPIA, instrumento similar ao RIPD no Brasil, quando o processamento de dados pessoais puder resultar em riscos altos de dano para os direitos e as liberdades individuais. Considerada a imprecisão legislativa para sua operacionalização, criou-se o grupo de trabalho denominado *Working Party 29 (WP29)* que elaborou o *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679 (Article 29 Data Protection Working Party, 2017)* dando maior clareza e aplicabilidade à norma. Embora as proposições do WP29 não sejam mais mandatórias dentro do UK, o ICO reforça que poderão ser referências úteis no processo de decisão de se proceder a uma avaliação de impacto.

Sem adentrar em detalhes específicos da GDPR, fica claro que a elaboração do DPIA não deve ser compreendida como mero cumprimento de conformidade, pois ele é um dos instrumento no qual os atores com responsabilidades sobre o processamento de dados pessoais não somente demonstram a conformidade com as normas e direitos, como também adotam uma cultura consciente, com decisões rastreáveis e responsáveis para a adoção de uma abordagem baseada em riscos, visando reduzir potenciais danos que esse tratamento possa acarretar aos proprietários dos dados.

O guia de avaliação de impacto propõe, em orientação às autoridades supervisoras de proteção de dados (DPAs) e às partes reguladas, a relação de 9 (nove) critérios que devem ser verificados para que auxiliem a identificação de quais tratamentos devem ser entendidos como associados a possíveis gerações de riscos altos. Quanto mais critérios forem aplicáveis, mais provável será que esteja lidando com a exigência de elaboração do DPIA, porém sem que apenas o enquadramento por um critério já não seja suficiente para caracterizar a necessidade de elaboração do DPIA. Estabelece, ainda, a obrigação

para que as DPAs tornem públicas a lista de operações de processamento que requerem a elaboração do DPIA.

O guia reforça a orientação de que a elaboração do DPIA seja anterior ao processamento, incentivando e fortalecendo a estratégia de desenvolvimento de sistemas *privacy by design*. Deve-se observar que o propósito e as conformidades da elaboração do DPIA, em geral, conduzem-no a um processo contínuo, em especial, quando a operação de processamento é dinâmica e sujeita a mudanças contínuas.

A GDPR inovou ao exigir, para alguns casos, a necessidade de estabelecimento pela organização de um responsável pela implementação da política de proteção dos dados pessoais na organização (*Data Protection Officer — DPO*). Como se vislumbra que em determinados momentos possa ocorrer alguma espécie de conflito de interesses entre o DPO e alguma liderança de área de negócio que esteja à frente de algum projeto, há a previsão na GDPR de que o DPO seja consultado na elaboração do DPIA e possa realizar sugestões, exigindo-se que esse processo para a tomada de decisão seja documentado no próprio instrumento. O guia ainda aborda outras questões importantes relacionadas ao processo estabelecido pela GDPR, como publicidade do DPIA e quando a consulta à DPA faz-se necessária.

O ICO lançou, em 2018, o Referencial *Data protection impact assessments* (ICO, 2018) para orientar os agentes regulados acerca do atendimento da GDPR. O documento inclui uma listagem não exaustiva e definitiva de exemplos apresentados, no qual a DPA do UK entende como necessária a realização do DPIA — conhecida como *blacklist* — complementando os critérios de exigência de DPIA sugeridos pelo WP29.

O *Government Digital Service (GDS)* e o *Office for Artificial Intelligence (OAI)*, com a participação do *The Alan Turing*, disponibilizaram, em 2019, diversos referenciais voltados para guiar o uso de IA pelo setor público que formam uma coleção rica para consulta dos diversos órgãos e servidores públicos: *Understanding artificial intelligence, Assessing if artificial intelligence is the right solution, Planning and preparing for artificial intelligence implementation,*

Managing your artificial intelligence Project, Understanding artificial intelligence ethics and safety.

O compartilhamento de bases de dados administrativos dentro da Administração Pública, que normalmente envolve dados pessoais, é peça fundamental para o avanço na implementação de sistema de IA no setor público. Visando avaliar as barreiras técnicas e legais sobre esse importante pilar, o *Centre for Data Ethics and Innovation (CDEI)* publicou, em 2020, o relatório *Addressing trust in public sector data use* (CDEI, 2020). Dentre os apontamentos, além da identificação de barreiras técnicas e culturais, destaca-se a opinião do risco de o ambiente confuso e complexo propiciar que interpretações e aplicações das determinações legais inconsistentes para o uso e o compartilhamento de dados pessoais possam prejudicar a confiança da sociedade no uso da tecnologia no setor público.

O GDS publicou, em 2020, o guia *Data Ethics Framework* voltado para guiar os servidores públicos para um uso apropriado, responsável e ético dos dados dentro da Administração. Estrutura-se em princípios e ações voltadas a promover a transparência, a responsabilidade e a justiça nos projetos de dados.

O ICO publicou, em 2020, o referencial *Guidance on IA and Data Protection* (ICO, 2020) com o objetivo de orientar as organizações a mitigar os riscos especificamente decorrentes de uma perspectiva de proteção de dados, explicando como princípios de proteção de dados devam ser aplicados a projetos de IA sem prejuízo à inovação.

Em que pese todo o esforço de diversas iniciativas não regulatórias das agências, um Comitê que assessora o Parlamento (EVANS, 2020) entende que sozinhas essas não têm sido suficientes para garantir um uso de IA responsável pelo setor público, tendo sido evidenciadas “deficiências notáveis”. Defende que o UK não precisa de um regulador para IA, mas sim que todos os reguladores devam preparar-se para enfrentarem os novos riscos que a IA poderá incorporar aos seus setores, sugerindo ainda que o CDEI venha a desempenhar um papel de centro de referência para apoiar os demais na temática. Pelo entendimento de que a implementação de sistemas de IA exige abordagem

baseada em riscos para uma adequada governança, faz-se necessária uma definição clara dos princípios éticos que devem ser integrados ao processo de governança de desenvolvimento de IA. Acrescente que considerando que a literatura nos últimos anos já indica a proposição de mais de 70 (setenta) princípios éticos, não ter uma definição clara de princípios que devam ser perseguidos poderá ocasionar a seleção casuística de alguns sem que isso garanta a indução do comportamento ético socialmente desejável.

A UE propôs, em abril de 2021, o seu primeiro referencial legal (EUROPEAN COMMISSION, 2021) visando unificar o endereçamento de riscos pelos diversos países do bloco no contexto de uso de IA. O referencial regulatório busca harmonizar os requisitos comuns obrigatórios ao projeto e ao desenvolvimento de determinados sistemas de IA antes de chegarem ao mercado, bem como busca padronizar os controles *ex-post* que serão conduzidos após a colocação em uso. Busca, ainda, estabelecer multas a responsáveis públicos e privados pelos descumprimentos de obrigações estabelecidas ao desenvolver ou usar IA que constitua alto risco para a segurança ou para os direitos fundamentais do cidadão. Também propõe que os responsáveis pelos sistemas de IA tenham que reportar às autoridades nacionais competentes a ocorrência de incidentes graves ou mau funcionamento que constituam violação das obrigações de direitos fundamentais, cabendo a essas autoridades a diligência de investigar, e que os países membros tenham autoridade nacional de supervisão, ainda que possam ter mais de uma autoridade nacional competente para tratar do tema. A proposta de regulação também define regras para identificar o uso de IA com probabilidade de alto risco de danos, devendo essa classificação ser feita com o objetivo pretendido do uso de IA no produto ou no serviço, e não na função isolada desempenhada pela IA, isto é, com foco na finalidade específica. Traz lista de IA não exaustiva de alto risco declarando que deverá ser periodicamente atualizada.

Desde sua saída da UE, notícias e proposições recentes de regulação sugerem que o UK venha a adotar posicionamentos regulatórios diversos. Recentemente, em julho deste ano, foi apresentada ao Parlamento pelo DCMS uma proposta chamada de *AI Regulation Policy Paper* (DCMS, 2022) definida

como sendo uma “abordagem clara, favorável à inovação e flexível para regular IA”. A proposta estabelece, dentre outros pontos, o estabelecimento de um conjunto de princípios não mandatórios num primeiro momento nos quais os reguladores setoriais deverão interpretar, priorizar e implementar dentro dos seus contextos específicos, visando criar uma regulação pró-inovação, proporcional aos riscos e adaptável.

Diferencia-se da proposta em discussão na UE em diversos pontos. A proposta apresentada no UK tende fortemente para construção de um ambiente regulatório voltado para a atração de desenvolvedores de IA para o país, por meio de proposição de um modelo de regulação descentralizado e com a moderação do atendimento dos princípios a ser feita pelo regulador do setor. Diferentemente da proposta do bloco europeu, não se propõe criar uma definição universal do que seria IA, a qual a regulação se aplicaria, não prevê a existência de uma autoridade que supervisione e harmonize os tratamentos de dados pelos diversos reguladores sobre o tema nem propõe a imposição de multas num eventual descumprimento da regulação. Reforça a diretriz para que os reguladores atuem principalmente por meio de orientações e guias e não especifica relação de usos vedados de sistemas de IA como proposto na UE. Há críticas (AKIN GUMP STRAUSS HAUER; FELD LLP, 2022) apontando que o desenho proposto possa trazer sobreposição de reguladores, dúvidas e confusão para consumidores e empresas de qual regulador seria o competente, inconsistência entre os poderes dos reguladores e lacunas na regulamentação existente.

A experiência do passado na introdução de outras tecnologias na sociedade moderna de produtos e serviços, no qual, em determinados segmentos, após um período inicial de discussões de como regular, passou-se para uma fase de exigências regulatórias mais rigorosas com exigências de serviços de auditoria ou de assecuração, serviu como impulso para que o CDEI produzisse o referencial *The roadmap to an effective AI assurance ecosystem*. A iniciativa visa, considerado os insumos qualificados presentes no país, projetar uma visão para posicionar o UK em lugar de destaque numa futura demanda de serviços de assecuração dentro de uma projeção de cadeia global de serviços.

Dessa forma, observa-se que o UK começa a sinalizar a intenção de proposição de diretrizes mais gerais de alto nível para que as agências reguladoras — num desenho de regulação descentralizada, flexível e específica para cada setor — possam customizar de acordo com suas especificidades contextuais, porém também com forte orientação para que se priorize um modelo de regulação por meio de medidas não impositivas e taxativas. No estágio atual, o UK dispõe de um rico acervo de medidas não regulatórias voltadas especificamente para IA, diversos lócus técnicos para endereçar discussões aprofundadas sobre o tema, porém com o risco de sobreposição ou mesmo conflito de princípios recomendados, devido à atuação de diversos atores com papel regulatório. Similarmente ao Brasil, possui uma DPA definida e com medidas regulatórias com obrigatoriedades estabelecidas pelo DPA 2018, que de forma indireta regula as IA que façam uso de dados pessoais.

CONCLUSÃO

O contexto de adoção do avanço tecnológico da IA pela sociedade há de ser conjugado com o atendimento dos avanços legais na grande maioria dos países democráticos para assegurar direitos fundamentais a todos os seus cidadãos, sem condicionantes ou restrições sob ótica de gênero, raça ou etnia.

O grande desafio é projetar uma adequada agenda regulatória pelos Estados que não iniba a inovação e garanta que a ampliação do uso dessas novas tecnologias permita a obtenção de ganhos de eficiência e de produtividade, sem que cause injustiças ou despreze outras legislações e direitos assegurados aos seus cidadãos.

Se pelo lado decisório sem IA, o processo em si já guarda desafios a serem vencidos pelos avaliadores para mitigação de erros oriundos de vieses ou ruídos sempre presentes e inerentes ao processo de construção de evidências, o aumento do uso de IA sem os devidos processos de consciência, concepção

e uso, pode aumentar ou propagar essa disfunção, prejudicando as decisões de alocação de recursos públicos nas políticas públicas.

Na jornada de crescimento de maturidade da comunidade e das instituições voltadas para avaliação de políticas públicas, faz-se necessário conhecer os riscos advindos da adoção dessa nova tecnologia para que esses possam ser mitigados e, assim, as evidências produzidas não ampliem ou perpetuem danos indesejáveis contra determinados grupos de indivíduos.

Uma leitura do desenho atual de regulação do tema, tomando por base os países comparados, leva-se a concluir que todos os reguladores deverão enfrentar o desafio de compreender e se preparar para adequadamente endereçar a regulação do uso de IA no setor, o que por si já será um grande desafio da Administração Pública nacional, frente às complexidades e às necessidades de mão de obra qualificada e diversificada, já escassa no mercado de trabalho.

Certamente, uma atuação de liderança de órgãos centrais, como já o fez em certa medida o CNJ, poderá induzir a construção de arranjos regulatórios que incentivem o uso responsável e consciente de IA dentro das funções estatais.

A forte interconexão do tema de IA e proteção de dados pessoais, se por um lado traz uma oportunidade para que a ANPD possa atuar como regulador transversal sobre a questão, influenciando positivamente as discussões pelos demais reguladores sobre a tecnologia e seus riscos, por outro, a realidade imposta de carência de estruturação adequada aos seus desafios mostra-se como barreira. A competência da agência para regulamentar o RIPD, de forma ampla a todos os atores que processam dados pessoais com uso de IA, é importante oportunidade para mitigar riscos e não pode continuar a sofrer atrasos.

A ausência de padrões transversais para que as agências possam usar para dosar e aplicar aos casos concretos, somado à presença de grandes empresas globais em determinados setores, poderá trazer desequilíbrios numa atuação regulatória esperada por parte do Estado para assegurar direitos e coibir abusos que possam decorrer do uso de IA. A discussão de proposição

legislativa em pleno andamento por meio da comissão de juristas poderá contribuir com essa lacuna, porém sem que se espere — à semelhança da visão comparada realizada em outros países — que venha ser suficiente para dar clareza e efetividade ao desenvolvimento de IA ética e confiável no país.

Dentro da condução da estratégia da EBIA pelo MCTI precisa haver a urgência de clareza da necessária promoção de liderança adequada para envolver e conscientizar todos os reguladores acerca dos desafios em relação ao tema, visando ao planejamento para endereçar o assunto dentro de suas competências, pois os desafios de usar e regular o mercado exigem o desenvolvimento de capacidades internas em tema complexo e muito dinâmico e quanto mais se demora para atuar, maior o risco de danos à sociedade.

Se sob a ótica do setor privado, a ausência de regulação clara alarga as margens de apetite ao risco e ao potencial retorno que possam adotar, num dever agir orientado que guiam a atividade estatal, as organizações públicas devem antecipar-se endereçando decisões de governança adequadas para o uso interno de IA dentro de suas competências.

Acredita-se que, para as demais agências com atuação regulatória, a busca de conhecimentos pelas experiências de países, nos quais as discussões com adoção de medidas não regulatórias já está mais avançada, poderá servir como acelerador na busca de insumos necessários para as discussões internas, pois a demanda que venha a regular o uso de IA nos seus diversos contextos de uso far-se-á necessária num horizonte temporal não muito longo, em especial aos que atuam em temas sensíveis como saúde, mercado financeiro, mobilidade, concorrência, entre outros.

Ainda que algumas proposições de legislações em discussão no exterior possam vir a endereçar obrigatoriamente mais claras e objetivas para o uso de IA, não se espera que venham a inibir o avanço de expansão no seu uso, nem que sejam muito diferentes ao exigirem uma abordagem baseada em riscos nos quais, em casos de uso de IA com alto risco, sejam exigidos processos mais sistematizados de avaliação de impacto, bem como de governança. Portanto, não se vislumbra que as discussões, em andamento no Congresso

Nacional, sigam caminho diverso aos desenhos que vêm sendo sugeridos no exterior, oscilando entre proposições mais restritivas como na proposta europeia – com proposição de relações de IAs proibidas ou mais principiológicas e customizáveis aos casos concretos, como em discussão no UK.

A ausência de atuação de um regulador externo não impede que as organizações públicas, inclusas as que produzam evidências para as tomadas de decisão voltadas para políticas públicas em qualquer de suas fases, dentro de seu processo de governança, estabeleçam diretrizes e proponham instâncias internas necessárias para que as boas práticas já discutidas em outros países na implementação de sistemas éticos de IA possam ser incorporadas no uso da IA. Essa ampliação do olhar da governança é fundamental para o fortalecimento de sua credibilidade decorrente da qualidade dos serviços prestados, do respeito às liberdades e aos direitos fundamentais do público-alvo, bem como da preservação dos valores organizacionais na consecução de seus objetivos. A exploração do potencial que a tecnologia permite precisa ser consciente, responsável e ético de forma a propiciar a melhoria dos serviços públicos entregues, o fortalecimento do movimento de políticas públicas baseadas em evidência e a contribuição com o aumento da confiança no Estado pela sociedade.

Considerada a rica disponibilidade de materiais técnicos já propostos, sugere-se que as áreas de governança das organizações públicas que fazem uso de IA comecem a endereçar os novos riscos impostos pelo uso da tecnologia por meio da adequação de seus gerenciamentos de riscos corporativos. Assim, os guias desenvolvidos, mesmo em outros países, sobre o tema, que em geral preconizam medidas de fortalecimento da governança e gerenciamento de riscos, já podem ser utilizados pelas organizações que queiram se antecipar a eventuais exigências regulatórias, ou mesmo, garantir que as IAs utilizadas estejam alinhadas a princípios legais e aos valores organizacionais de governança pública.

Como descrito em um referencial de Singapura (SG:D., 2020), é importante que para eventuais revisores ou utilizadores de informações de um sistema de IA que ela não produz fato, isto é certeza, mas sim probabilidade de determinada

resposta a um problema que lhe seja dado como objetivo. Compreender e fomentar essa consciência, reduzindo o risco de complacência humana é um ponto relevante no processo decisório supervisionado por humanos, o que, em geral, são as utilizações para tomadas de decisão em políticas públicas.

Compreendendo a complexidade da geração de evidências e o seu uso pelos tomadores de decisão, fica claro que é fundamental o aumento de consciência por todas as partes envolvidas no processo dos riscos e das limitações decorrente do uso de IA, para que seja possível monetizar os ganhos esperados pelo seu uso para a sociedade, sem que na prática se provoque danos a direitos assegurados. A ampliação de seu uso como ferramenta de decisão ou apoio a essa, nas diversas finalidades, quer dentro de processos internos, quer para a prestação de serviços públicos pelo Estado, é uma grande oportunidade de reduzir a variabilidade em decisões decorrentes de ruídos nas decisões humanas, porém desde que seja concebida dentro de uma adequada governança de forma sistemática e criteriosa, consideradas as diversas fases e fontes na qual está sujeita a introdução de erros por vieses ao longo do seu ciclo de vida.

REFERÊNCIAS

AKIN GUMP STRAUSS HAUER & FELD LLP (org.). **The UK Government's Early Proposals for Innovation-Friendly AI Regulation**. 2022.

BAROCAS, S. *et al.* **Fairness and machine learning: limitations and opportunities**. [S. L.]: Fairmlbook, 2019.

BRASIL. **Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre inteligência artificial no Brasil - CJSUBIA**, Subsidiar a elaboração de minuta de substitutivo com objetivo de estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil, 2022a.

BRASIL. **Lei n. 12.527, de 18 de novembro de 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no §

2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências, 2011.

BRASIL. **Lei n. 13.709, de 14 de agosto de 2018.** Lei Geral De Proteção De Dados Pessoais - LGPD. Diário Oficial da União, Edição extra, Brasília, DF, 15 de ago 2018.

BRASIL. Tribunal de Contas da União (Plenário). **Acórdão n. 1.139/2022.** Processo n. 006.662/2021-8. Levantamento de Auditoria Inteligência Artificial. Relator: Aroldo Cedraz, de 25 de mai de 2022b.

BRASIL. Tribunal de Contas da União (Plenário). **Acórdão n. 1.384/2022.** Processo n. 039.606/2020-1. Auditoria para avaliar as ações governamentais e os riscos à proteção de dados pessoais. Relator: Ministro Augusto Nardes, de 15 de junho de 2022c.

CENTRE FOR DATA ETHICS AND INNOVATION (CDEI). **Addressing trust in public sector data use.** Reino Unido, 2020.

CHESTERMAN, S. **We, the robots?:** regulating artificial intelligence and the limits of the law. Cambridge: Cambridge University Press, 2021. 289 p.

COMITÊ CENTRAL DE GOVERNANÇA DE DADOS. **Guia de Boas Práticas:** Lei geral de proteção de dados (LGPD). BRASIL, 2020. 69 p.

COMMITTEE ON STANDARDS IN PUBLIC LIFE. **The Seven Principles of Public Life** - NOLAN Principles. 31 de maio de 1995, 4 p. 1995.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Poder Judiciário Conselho Nacional de Justiça, RESOLUÇÃO 332/2020,** 2020.

DOMANSKI, R. *et al.* Toward an ethics of digital government: A first discussion. **ACM International Conference Proceeding Series,** 2018.

ENGLER, A. **New White House guidance downplays important AI harms.** Disponível em: <<https://www.brookings.edu/blog/techtank/2020/12/08/>>

new-white-house-guidance-downplays-important-ai-harms/>. Acesso em: 15 ago. 2022.

ENGSTROM, D. F. *et al.* **Government by algorithm: Artificial Intelligence in Federal Administrative Agencies** NNYU School of Law. [s.l.: s.n.]. Disponível em: <https://www.law.ox.ac.uk/sites/files/oxlaw/government_by_algorithm_acus_report.pdf>. Acesso em: 10 ago. 2022.

ESTADOS UNIDOS DA AMÉRICA (EUA). **Executive Order 13.859, de 11 de fev. de 2019.** Maintaining American Leadership in Artificial Intelligence.

ESTADOS UNIDOS DA AMÉRICA (EUA). **Executive Order 13.960, de 3 de dez. de 2020.** Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government.

EUROPEAN COMMISSION (EC). **Article 29 Data Protection Working Party WP248 rev. 1**, de 04 de abr. de 2017. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679.

EUROPEAN COMMISSION (EC) - GPAN AI. **Orientações Éticas para uma IA de Confiança, de 8 de abr. de 2019. BÉLGICA.**

EUROPEAN COMMISSION (EC). **Proposal for a Regulation laying down harmonised rules on artificial intelligence**, 2021.

EVANS, L. **Artificial Intelligence and Public Standards A Review by the Committee on Standards in Public Life.** Reino Unido, 2020.

FEW, S. **The Data Loom: Weaving Understanding by Thinking Critically and Scientifically with Data.** CA. Analytics Press, 2019. 133 p.

FRENCH, O. *et al.* **bias (n .)**. Disponível em:<<https://www.etymonline.com/word/bias>> . Acesso em: 25 maio 2022.

GOVERNMENTAL DIGITAL SERVICE: OFFICE FOR ARTIFICIAL INTELLIGENCE. **A guide to using artificial intelligence in the public sector.** UK, 2019. 48 p.

INFORMATION COMMISSIONER'S OFFICE (ICO), **Project Explain. Explaining decisions made with AI Draft-Guidance Part 2: Explaining AI in practice.** Reino Unido, 2019.

INFORMATION COMMISSIONER'S OFFICE (ICO). **Accountability and governance Data Protection Impact Assessments (DPIAs)**, 2018.

INFORMATION COMMISSIONER'S OFFICE (ICO). **DPIA template**, 2018. Disponível em: <<https://ico.org.uk/media/2258461/dpia-template-v04-post-comms-review-20180308.pdf>>. Acesso em: 18 ago. 2022.

INFORMATION COMMISSIONER'S OFFICE (ICO). **Guide to the UK General Data Protection Regulation (UK GDPR)**.

INFORMATION COMMISSIONER'S OFFICE (ICO). **Guidance on AI and Data Protection**, 2020.

INFORMATION COMMISSIONER'S OFFICE (ICO). **AI and Data Protection Risk Toolkit**.

KAHNEMAN, D. et al. **Noise: a flaw in human judgment**. London: William Collins, 2021.

KAHNEMAN, D. **Rápido e devagar: duas formas de pensar**. Rio de Janeiro. Editora Objetiva, 2011.

KARL POPPER. Disponível em: <https://pt.wikipedia.org/wiki/Karl_Popper>. Acesso em: 14 ago. 2022.

MALEK, M. A. Criminal courts' artificial intelligence: the way it reinforces bias and discrimination. **AI and Ethics**, v. 2, n. 1, p. 233–245, fev. 2022.

MARTIN, K. **Algorithmic Bias and Corporate Responsibility: How companies hide behind the false veil of the technological imperative** (August 14, 2021). Ethics of Data and Analytics. Kirsten Martin (Ed.). Taylor & Francis.

MEHRABI, N. *et al.* A Survey on Bias and Fairness in Machine Learning. **ACM Computing Surveys**, v. 54, n. 6, 2021.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES (MCTI). **Portaria n. 4.617, de 06 de abril de 2021.** Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília, DF: Diário Oficial da União, 12 abr. 2021. n. 67, Seção 1, p. 30.

NARAYANAN, A. **FAT* tutorial: 21 fairness definitions and their politics.** New York: 2018. Disponível em: <https://www.youtube.com/watch?v=jlXluYdnyyk>. Acesso em: 18 ago. 2022.

NTOUTSI, E. *et al.* **Bias in Data-driven AI Systems - An Introductory Survey.** arXiv, 2020.

OCDE. **Open Government Data Report - Enhancing Policy Maturity for Sustainable Impact.** 2018.

OFFICE FOR ARTIFICIAL INTELLIGENCE (OAI). **Guidelines for AI procurement,** 2020.

PAUL, R.; ELDER, L. **Critical Thinking: Tools for Taking Charge of Your Professional and Personal Life.** [s.l: [s.n.].

SECRETARIA DE GOVERNO DIGITAL (SGD/ME). **Guia e template do Relatório de Impacto à Proteção de Dados Pessoais,** 2020.

SG:D. **Model Artificial Intelligence Governance Framework Second edition.** Singapore, 2020.

SUSSMAN, H.; MCKENNEY, R.; WOLFINGTON, A. U . S . **Artificial Intelligence Regulation Takes Shape.** Disponível em: <<https://www.orrick.com/en/Insights/2021/11/US-Artificial-Intelligence-Regulation-Takes-Shape>>. Acesso em: 25 ago. 2022.

THE ALAN TURING INSTITUTE. **Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector.** arXiv Computer Science, 2019.

TRANSPARÊNCIA BRASIL. **Estrutura de Avaliação de Riscos a Direitos e de Transparência - Uso de Inteligência Artificial pelo Poder Público,** 2020.

TRANSPARÊNCIA BRASIL. **Recomendações de Governança - Uso de Inteligência Artificial pelo Poder Público**. Brasil, 2020.

U.S. GOVERNMENT ACCOUNTABILITY OFFICE (GAO). **Framework for Federal Agencies and Other Entities**, 2021.

ZAMRODAH, Y. **Diretrizes de auditabilidade e conformidade no desenvolvimento e testes de soluções de ia no âmbito do LIAA-3R** - 2ª edição. v. 15, n. 2, p. 1-23, 2022.

ZOOK, M. *et al.* Ten simple rules for responsible big data research. **PLoS Computational Biology**, v. 13, n. 3, p. 1-10, 2017.



TEMA: ENSINO SUPERIOR

FERRAMENTA:
APRENDIZADO DE MÁQUINA
SUPERVISIONADO

MACHINE LEARNING APLICADA À ANÁLISE DOS FATORES DETERMINANTES DA TAXA DE OCUPAÇÃO E CONCLUSÃO NO ENSINO SUPERIOR NO BRASIL¹

ISVALDO SILVEIRA LIMA FILHO⁽⁷⁾

INTRODUÇÃO

A relevância de políticas voltadas à promoção do Ensino Superior (ES) decorre do papel desse nível de ensino em potencializar o aumento de produtividade e competitividade à economia nacional, mas também, e não menos importante, de proporcionar meios para promover maior inclusão e melhores condições de vida a grupos sociais mais carentes. No Brasil, essa importância fica ainda mais evidente quando observamos a evolução das taxas de matrícula no Ensino Superior, especialmente quando comparadas as de outros países. De fato, estamos atrás da maioria dos países latino-americanos nesse quesito (HERINGER, 2018). A Pesquisa Nacional por Amostra de Domicílios (PNAD) de

¹ Este artigo resulta de trabalho de conclusão orientado pelo Professor James Matos Sampaio.

2019 mostrou que a taxa de matrícula no Ensino Superior (entre jovens de 18 a 24 anos) era de 20,6% (IBGE, 2019), pequena, se comparada a meta de atingir 33% em 2024 — PNE 2014-2024 (BRASIL, 2014).

Entretanto, em que pese os meritórios esforços para a ampliação de vagas no ES, esses contrapõem-se à constatação de que parte substancial dessas vagas ficam ociosas, seja por insuficiência do quantitativo de ingressantes vis à vis ao quantitativo ofertado, seja pela evasão de parte dos alunos no transcorrer dos cursos, o que importa em desperdício de recursos públicos e privados. Isso evidencia a importância de uma análise que contemple não apenas a evasão ou a taxa de sucesso dos cursos, mas que considere a ociosidade das vagas inicialmente ofertadas.

Por outro lado, nesse contexto de avaliação de políticas públicas, tem ganhado relevância a aplicação de técnicas de Aprendizado de Máquina (*Machine Learning*), as quais permitem o tratamento e a inspeção de grandes conjuntos de dados, identificando padrões e características que permitem análise de fenômenos complexos, influenciados por uma gama de variáveis interrelacionadas (HASTIE; TIBSHIRANI; FRIEDMAN, 2008).

Diante disso, o objetivo deste trabalho foi investigar e mensurar, por meio de técnicas de *Machine Learning* (ML), os principais fatores que contribuem para a taxa de ocupação e conclusão dos cursos de graduação, considerando tanto características inerentes às localidades de ofertas dos cursos, quanto aquelas relativas às instituições, aos cursos e ao perfil dos ingressantes, de modo a explorar o potencial das técnicas de ML na construção de modelos que possam subsidiar os gestores na avaliação dessas políticas, na busca por torná-las mais eficazes.

APRENDIZADO DE MÁQUINA (*MACHINE LEARNING*)

Hastie, Tibshirani e Friedman (2008) referem-se à ciência do aprendizado como aquela que desempenha um importante papel nos campos da estatística e da computação (mineração de dados e inteligência artificial). Tipicamente,

como mencionam esses autores, busca-se a predição para uma determinada medida (quantitativa ou categórica – variável dependente) baseado em um conjunto de *features* (preditores ou variáveis independentes). Em vez de especificar uma série de regras (por meio de uma linguagem de programação), a máquina (ou o algoritmo) teria como apreendê-las a partir da observação de um conjunto de dados, identificando padrões que determinem, com certa acurácia, os resultados pretendidos, com aplicações especialmente com objetivo de predição e otimização (QIFANG BI et al. 2019).

BASES DE DADOS E ATRIBUTOS UTILIZADOS

Neste trabalho, foi utilizada como fonte principal de dados os microdados do Censo Superior disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O ano selecionado foi o de 2019, por tratar-se do censo mais recente, anterior ao evento da pandemia da covid-19, de modo a evitar distorções de análises.

O objeto de interesse de análise recaiu sobre cursos ativos com concluintes (cursos recentes não foram considerados), bem como cursos nos quais o número de matriculados superou o número de ingressantes (cursos já consolidados). As estimativas populacionais adotadas também foram com base nesse mesmo ano (IBGE, 2021), e foi utilizado o conceito de Regiões Geográficas Imediatas (RGI) para tratamento dos dados das localidades de oferta dos cursos (IBGE, 2017).

Foram selecionadas, considerando a disponibilidade e as causas da evasão apontadas em pesquisas de referência (SALES JÚNIOR, 2013), as seguintes variáveis preditoras: 1) duração do curso (em anos); 2) grau acadêmico (bacharelado/licenciatura); 3) polo (se curso é ofertado em município, polo ou não); 4) Sexo (proporção de matriculados do sexo feminino); 5) Cota (proporção de cotistas dentre os matriculados); 6) Turno (proporção de matriculados no turno noturno); 7) Financiamento (proporção de matriculados que utilizam alguma fonte de financiamento); 8) Assistência estudantil (proporção de matriculados beneficiários de programa de assistência estudantil); 9) Faixa

etária ingressante (*rank* de 2 a 9 conforme proporção de ingressantes em cada faixa etária); 10) Concorrência (inscritos por vagas ofertadas); 11) Ingressantes Enem (proporção de ingressantes via Enem); 12) Origem (proporção de matriculados oriundos de escola pública); 13) Atividade extracurricular (proporção de alunos envolvidos em atividades extra); 14) ENADE-CPC (faixa de 1 a 5); 15) Critério de escolha da IES (proporção de concluintes que escolheram a IES por critérios de qualidade/prestígio); 16) Rendimento mensal per capita da RGI (valor mediano); 17) Faixa populacional da RGI de oferta do curso; 18) Matrículas no EM na RGI (proporção em relação à população); 20) Total Candidatos do Enem na RGI (proporção em relação à população); 21) Notas (média) do Enem na RGI.

Foi, ainda, analisada a existência de correlação entre as variáveis numéricas e, por meio de inspeção visual (gráfica), identificada correlação significativa apenas entre as variáveis de Rendimento Per Capta (por RGI) e a Média da Notas dos candidatos do Enem (também por RGI), sendo considerada para o modelo final a variável referente ao Enem, uma vez que os dados de rendimentos foram obtidos com um lapso temporal maior (Censo 2010) (Correlação Spearman = 0.75 p-value \approx 0).

VARIÁVEIS DE INTERESSE (*TARGETS*)

Como variáveis de interesse (*targets*) foram escolhidas as seguintes:

- 1) Taxa de Ocupação = $Qt \text{ Matrículas} / (Qt \text{ vagas anuais} * \text{Dur Curso})$: A taxa de ocupação reflete o interesse dos alunos pelo curso e a permanência até sua conclusão.
- 2) Taxa de Conclusão = $Qt \text{ Concluintes} / Qt \text{ vagas anuais}$: A taxa de conclusão dá uma medida de sucesso do curso, porém nota-se que não corresponde precisamente à taxa de sucesso como definida pelo MEC.

ESCOLHA DOS ALGORITMOS (ESTIMADORES)

Fez-se a escolha dos algoritmos a serem utilizados por meio do pacote *Lazy Predict*, que tem o objetivo de auxiliar a construir e a testar uma série de modelos básicos (sem “*tuning*” dos hiperparâmetros) para vários algoritmos (estimadores) (PANDALA, 2020). Os estimadores selecionados foram: *XGBClassifier* (XGB), *LGBMClassifier* (LGBM), *ExtraTreesClassifier* (XTrees) e o *RandomForestClassifier* (RF).

As medidas de desempenho dos estimadores utilizadas neste trabalho foram (FERNANDES, 2021): *acurácia* — mede o percentual de acertos geral (total de acertos/total de elementos testados); *precisão* — mede o percentual de verdadeiros positivos sobre o total de valores identificados como positivos; *recall* (sensibilidade): percentual de acerto na identificação dos casos positivos testados; *f1 Score*: média harmônica calculada com base na *precisão* e no *recall*.

RESULTADOS E DISCUSSÃO

A seguir são apresentados os resultados obtidos para cada subgrupo analisado, bem como análise das variáveis com maior influência na aplicação dos modelos obtidos a partir dos algoritmos selecionados. Apesar de não permitir interpretação válida de relação de causalidade, essa análise permite verificar as variáveis com maior “peso” na resolução dos modelos utilizados, indicando possíveis implicações para os tomadores de decisão das políticas públicas envolvidas no objeto em estudo.

1. TAXA DE OCUPAÇÃO

1.1 Taxa de Ocupação — IES Públicas, cursos presenciais (bacharelado e licenciatura).

Para esse subgrupo, obteve-se o seguinte desempenho dos estimadores utilizados:

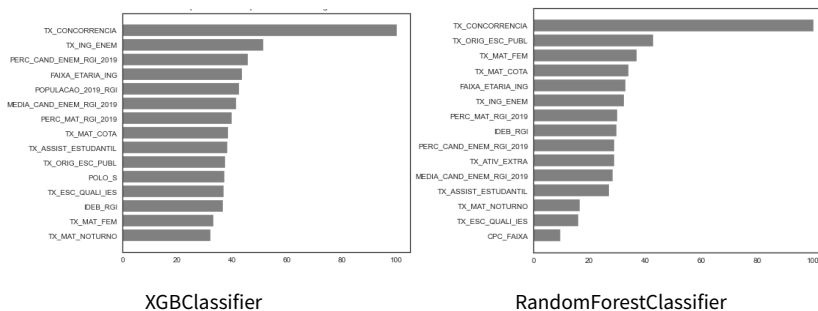
Quadro 1 | Medidas desempenho — item 1.1

Estimador	Acurácia	Precisão	Recall	F1 Score
XGBClassifier	0,71	0,68	0,71	0,68
LGBMClassifier	0,70	0,68	0,70	0,69
ExtraTreesClassifier	0,69	0,66	0,69	0,67
RandomForestClassifier	0,71	0,69	0,71	0,69

Fonte: Elaborado pelo autor.

Apresenta-se, nos quadros abaixo, a contribuição de cada variável (*feature*) para o desfecho calculado (*target*) dos dois modelos de melhor desempenho.

Gráfico 1 | Taxa de Ocupação — IES Públicas — item 1.1



Fonte: Elaborado pelo autor.

Em termos de importância das *features*, os modelos mostraram-se convergentes, à exceção do XGBooster que apresentou diferenças relevantes para cerca de 25% das variáveis utilizadas. A mais notável diferença diz respeito à população da região (RGI) na qual o curso é ofertado. Nos outros modelos, não aparece como variável relevante, porém, nesse, está entre as 5 (cinco) mais

influentes. É razoável supor que cursos oferecidos por IES públicas em grandes centros tenham uma qualidade e, por conseguinte, uma taxa de ocupação maior, inclusive porque historicamente esses cursos foram estruturados inicialmente nessas localidades, contando, portanto, com longa tradição e maior reconhecimento. Além disso, também é esperada uma demanda maior por Ensino Superior nessas localidades, o que também influencia positivamente.

Conforme gráfico 1, pode-se observar que as variáveis com maiores contribuições para taxa de ocupação, segundo os modelos utilizados, foram, na ordem de importância (mais importante para a menos importante), as seguintes:

- a) TX CONCORRÊNCIA — Era de se esperar tal resultado, uma vez que quanto maior o interesse de ingresso pelo curso, maior seu prestígio social, sendo razoável esperar um comprometimento maior dos alunos em sua permanência. Também pode refletir um nível de formação básica melhor do alunado resultante de uma competitividade maior no ingresso. Ressalte-se que sempre foi a variável de maior importância em todos os 4 (quatro) modelos utilizados.
- b) TX ORIGEM ESC PUBL — Cursos com maior participação de alunos oriundos de escola pública tendem a ter pior resultado em termos de desempenho, o que pode ser explicado a partir da qualidade deficiente da educação básica pública no Brasil. Essa variável só não apareceu como tão relevante no modelo do XGB.
- c) TX MAT FEMININA — Cursos com participação maior de estudantes do sexo feminino têm um desempenho melhor. Novamente, essa variável só não se apresentou com tanta relevância no modelo do XGB.
- d) TX COTA — Proporção maior de cotistas tem efeito positivo em relação às taxas de ocupação. Nota-se bem que a política de cotas parece não prejudicar as taxas de ocupação, ao contrário, é fator que influencia positivamente no desempenho dos cursos.
- e) FAIXA ETÁRIA ING – Cursos com estudantes de menor faixa etária no momento do ingresso também têm, em geral, uma taxa de ocupação maior. Alunos “fora de faixa” podem estar associados, de forma mais frequente, a outras atividades e a responsabilidades sociais (como chefes de famí-

- lia) que trazem desafios extras para o acompanhamento de um curso de graduação. Essa influência só não foi captada (nesse nível de importância) pelo modelo do LGBM.
- f) TX ING ENEM — cursos com maior proporção de ingressantes via Enem tendem a ter um melhor desempenho. A utilização da nota do Enem pode apontar para uma seleção mais rigorosa/criteriosa dos ingressantes quando comparada aos processos seletivos próprios. Esse fator mostrou-se relevante em todos os modelos empregados (nos modelos do LGBM e do XGB foram respectivamente o 3º e o 2º fator mais influentes).
- g) Outros fatores importantes — fatores ligados à qualidade e ao nível de oferta do Ensino Médio na região do curso (IDEB, proporção de matrículas no Ensino Médio, proporção de candidatos e média de notas no Enem), à proporção de estudantes universitários envolvidos em atividades extracurriculares e atendidos em programas de assistência estudantil. As médias do Enem, inclusive, foram fatores dos mais influentes nos modelos do LGBM e XGB. A participação em atividade extra, por sua vez, não contou muito para o resultado do modelo do XGB.

É de se notar ainda que fatores como a utilização de financiamento (o que era esperado para IES públicas), a duração do curso (as variações de duração são pequenas para cursos de bacharelado e licenciaturas), o grau acadêmico (bacharelado ou licenciatura), a avaliação do curso (ENADE-CPC — em geral, os cursos das IES públicas são bem avaliados), o critério de escolha da IES e mesmo o turno tiveram pouca influência no resultado final.

Portanto, verifica-se que fatores institucionais não são tão relevantes no caso de IES públicas, provavelmente devido ao fato de elas já serem instituições mais consolidadas e respeitadas, além disso, são mais acessíveis (ensino público e gratuito). Porém, fatores que refletem o prestígio social do curso (concorrência, processo seletivo) e condições pessoais do aluno (sexo, faixa etária, formação do ensino básico), além de fatores ligados ao suporte e à integração acadêmicas (atividades extracurriculares e assistência estudantil) têm marcante influência sobre o desempenho desses cursos.

1.2 Taxa de Ocupação – IES Privadas, cursos presenciais (bacharelado e licenciatura)

Quadro 2 | Medidas desempenho — item 1.2

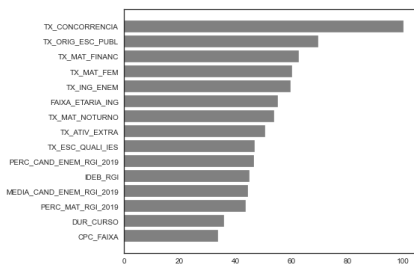
Estimador	Acurácia	Precisão	Recall	F1 Score
XGBClassifier	0,78	0,74	0,78	0,75
LGBMClassifier	0,79	0,74	0,79	0,75
ExtraTreesClassifier	0,79	0,74	0,79	0,74
RandomForestClassifier	0,78	0,71	0,78	0,71

Fonte: Elaborado pelo autor.

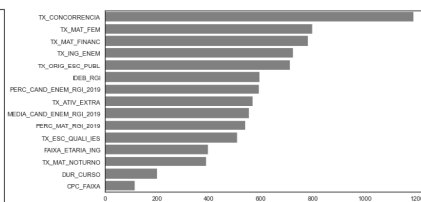
Observa-se que os modelos apresentaram-se com maior poder preditivo (quase 80% de acurácia) no caso das “Taxa de Ocupação” para as IES privadas.

Apresenta-se, no gráfico abaixo, a contribuição de cada variável (*feature*) para o desfecho calculado (*target*) nos modelos de maior desempenho.

Gráfico 2 | Taxa de Conclusão — IES Públicas — item 1.2



ExtraTreesClassifier



LGBMClassifier

Fonte: Elaborado pelo autor.

Primeiro, verifica-se que houve variação significativa (e esperada) da importância das variáveis para o resultado preditivo dos modelos, quando comparado com aqueles obtidos para as IES públicas.

Conforme gráfico 2, pode-se observar que as variáveis com maiores contribuições para taxa de ocupação, segundo os modelos utilizados, foram, na ordem de importância (mais importante para a menos importante), as seguintes:

- a) TX CONCORRÊNCIA — Também se mostrou mais influente nesse caso, o que reforça a assertiva de que quanto maior o interesse de ingresso pelo curso, maior seu prestígio social, sendo razoável esperar um comprometimento maior dos alunos em sua conclusão, bem como uma melhor formação dos ingressantes. Destaca-se, porém, que, para o modelo do XGB, a importância dessa *feature* é superada por aquela que se refere à duração do curso.
- b) TX ORIGEM ESC PUBL — Novamente, repete-se para o caso das IES privadas, semelhantemente ao notado para as IES Públicas.
- c) TX FINANCIAMENTO — Fator mais relevante em relação às IES privadas (por óbvio), quando se verifica que uma proporção maior de alunos com financiamento tem efeito positivo em relação às taxas de ocupação. Apenas no modelo do XGB é que essa variável não aparece entre as mais influentes.
- d) DUR CURSO — Novamente em contraste com as IES públicas, para as IES privadas, a duração do curso tem importante papel em suas taxas de ocupação, pode haver aí a influência de questões de financiamento e de necessidade de geração de renda por parte dos alunos. Isso ficou mais evidenciado nos modelos do RF e do XGB.
- e) TX MAT FEMININA — Também é um fator comum, entre IES públicas e privadas. Essa influência só não foi tão ressaltada no modelo do XGB.
- f) TX ING ENEM, PERC CAND ENEM e MÉDIA CAND ENEM — Para IES privadas também ocorre que cursos com maior proporção de ingressantes via Enem tendem a ter melhor desempenho. Novamente, a utilização da nota do Enem pode apontar para uma seleção mais rigorosa/criteriosa dos ingressantes quando comparada aos processos seletivos próprios. Além disso, nesse caso, apareceu também como importante o curso ser ofertado em

uma área com maior (e mais qualificada) demanda por Ensino Superior (mais participantes da prova do Enem na região, com maior nota média).

g) Outros fatores importantes — fatores ligados ao turno do curso, à participação em atividades extracurriculares (fator relevante principalmente no modelo XGB) e ainda aos critérios de escolha da IES (se a escolha é por prestígio/qualidade, o curso tende a ter um melhor desempenho).

É de se notar ainda que fatores como o tamanho da população da região de oferta (relevante apenas no modelo XGB), o grau acadêmico (bacharelado ou licenciatura), a avaliação do curso (ENADE-CPC), as cotas (ainda não está consolidada nas IES privadas uma política de cotas), a assistência estudantil (também ainda pouco presente na IES privada), além de faixa etária (relevante apenas no modelo XTrees) tiveram pouca influência no resultado final.

Portanto, verifica-se que fatores institucionais passam a ter mais relevância no caso de IES privadas (critérios de escolha das IES), enquanto que semelhantemente às IES públicas também são importantes fatores que refletem o prestígio social do curso (concorrência, processo seletivo) e condições pessoais do aluno (sexo e formação do ensino básico), porém a faixa etária e o rendimento (da região) não se apresentam mais como relevantes nesse segmento, mas sim o acesso ao crédito estudantil. A integração acadêmica (atividades extracurriculares) continua também sendo um fator de marcante influência sobre o desempenho (taxa de ocupação) dos cursos desse segmento.

2. TAXA DE CONCLUSÃO

2.1 Taxa de Conclusão — IES Públicas, cursos presenciais (bacharelado e licenciatura)

Quadro 3 | Medidas desempenho — item 2.1

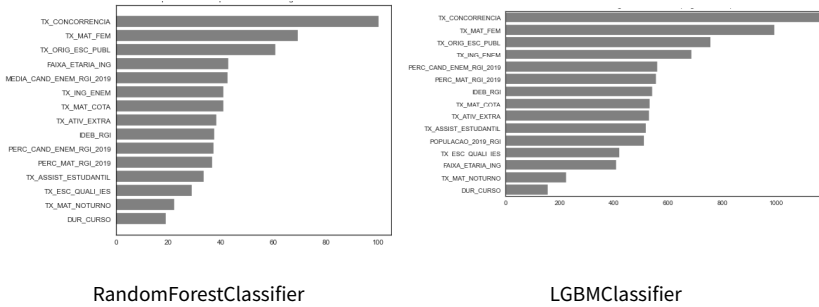
Estimador	Acurácia	Precisão	Recall	F1 Score
XGBClassifier	0,60	0,60	0,60	0,60
LGBMClassifier	0,62	0,61	0,62	0,61
ExtraTreesClassifier	0,59	0,58	0,59	0,58
RandomForestClassifier	0,62	0,62	0,62	0,61

Fonte: Elaborado pelo autor (2022).

Nota-se que o desempenho preditivo dos modelos cai consideravelmente para o caso da taxa de conclusão. É possível que outros fatores, talvez mais ligados à retenção, não estejam sendo devidamente contemplados. Informações como a nota de entrada dos alunos e a repetência também poderiam incrementar o desempenho desses modelos. Variações na quantidade de vagas ofertadas também podem distorcer o indicador (*target*) utilizado.

Apresenta-se, no gráfico a seguir, a contribuição de cada variável (*feature*) para o desfecho calculado (*target*) para os modelos de melhor desempenho.

Gráfico 3 | Tx conclusão — IES Públicas — item 2.1



Fonte: Elaborado pelo autor (2022).

O modelo do XGB foi novamente o que apresentou alguma variação mais marcante. Notavelmente, a duração do curso e o critério de escolha da instituição mostraram-se como importantes nesse caso.

A partir do gráfico 3, comparando-se os resultados para a Taxa de Conclusão com aqueles obtidos para Taxa de Ocupação (1.1) para IES públicas, verifica-se que os principais fatores permanecem os mesmos. Entretanto, o peso de outras variáveis, principalmente no que se refere a sexo e a origem (escola pública/privada), apresentaram aumento em sua importância em comparação com a variável taxa de concorrência (modelos mais equilibrados). Ademais, nota-se ainda, como diferença, uma pequena elevação na influência do fator relativo à nota do Enem, o que pode ser indicativo de que a nota de ingresso no curso (perceba que se está usando apenas uma *proxy*) tem importância maior quando se trata do desempenho em relação às taxas de conclusão.

Outros fatores importantes foram: a participação de alunos em atividades extracurriculares, a proporção de matrículas e a qualidade (IDEB) do Ensino Médio da região e a assistência estudantil.

Nota-se ainda que fatores como a utilização de financiamento (o que era esperado para IES públicas), o grau acadêmico (bacharelado ou licenciatura), a avaliação do curso (ENADE-CPC — em geral, os cursos das IES públicas são bem avaliados) e mesmo a duração do curso tiveram pouca influência no resultado final (com exceção do modelo XGB). Ressalta-se que esses mesmos

fatores também se mostraram de menor importância quando o *target* era a taxa de ocupação dos cursos em IES públicas.

Da mesma forma que analisado no item 1.1, aqui também se verifica que fatores institucionais não são tão relevantes no caso de IES públicas, que fatores refletem o prestígio social do curso (concorrência, processo seletivo) e as condições pessoais do aluno (sexo, faixa etária, formação do ensino básico), além de fatores ligados ao suporte e à integração acadêmicas (atividades extracurriculares e assistência estudantil) têm marcante influência sobre o desempenho desses cursos.

2.2 Taxa de Conclusão — IES Privadas, cursos presenciais (bacharelado e licenciatura)

Quadro 4 | Medidas desempenho — item 2.2

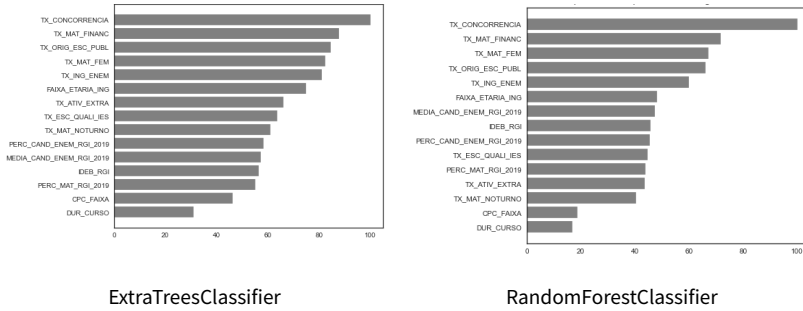
Estimador	Acurácia	Precisão	Recall	F1 Score
XGBClassifier	0,62	0,60	0,62	0,61
LGBMClassifier	0,64	0,61	0,64	0,62
ExtraTreesClassifier	0,65	0,63	0,65	0,63
RandomForestClassifier	0,67	0,64	0,67	0,63

Fonte: Elaborado pelo autor (2022).

No caso das IES privadas, os estimadores mostram-se com melhor desempenho para a Taxa de Conclusão, quando comparados com o mesmo *target* para IES públicas, especialmente o RF.

O gráfico a seguir, apresenta a contribuição de cada variável (*feature*) para o desfecho calculado (*target*) para os modelos de maior desempenho.

Gráfico 4 | Tx conclusão — IES Privadas — item 2.2



Fonte: Elaborado pelo autor (2022).

Novamente, o modelo XGB apresentou maiores variações em relação à importância das *features*, quando comparado aos demais estimadores. O fator populacional, por exemplo, só teve sua importância captada por esse modelo. Semelhantemente ao que ocorrera para a taxa de ocupação, verifica-se que houve também uma variação importante (e esperada) da importância das variáveis para o resultado preditivo do modelo em relação às IES públicas, especialmente com relação ao financiamento estudantil (relevante para as IES privadas) e à taxa de alunos cotistas (relevante para as IES públicas).

O gráfico 4, apresenta as variáveis com maiores contribuições para taxa de conclusão, segundo os modelos utilizados. Na ordem de importância (mais importante para a menos importante), foram as seguintes: TX CONCORRÊNCIA; TX FINANCIAMENTO; TX MAT FEMININA; TX ORIGEM ESC PUBL; TX ING ENEM; FAIXA ETÁRIA ING; MÉDIA NOTAS ENEM (conforme já comentado anteriormente — vide item 1.2).

Outros fatores importantes — fatores ligados à qualidade do Ensino Médio IDEB, proporção de candidatos do Enem e proporção de matrículas no Ensino Médio na região, participação dos universitários em atividades extracurriculares, além de motivação para escolha da instituição (fatores ligados ao prestígio/qualidade ou relacionados à conveniência/oportunidade).

Novamente, para as IES privadas, fatores como o tamanho da população da região de oferta (só apareceu como relevante no modelo do XGB), o grau

acadêmico (bacharelado ou licenciatura), a avaliação do curso (ENADE-CPC), as cotas (ainda não está consolidado nas IES privadas uma política de cotas) e mesmo a assistência estudantil (também ainda pouco presente na IES privada) tiveram pouca influência no resultado final.

Comparando-se os resultados para a Taxa de Conclusão com aqueles obtidos para Taxa de Ocupação (1.1) para IES privadas, verifica-se que os principais fatores permanecem praticamente os mesmos. Entretanto, o peso de outras variáveis, como ocorreu também na comparação para IES públicas (Tx Ocupação X Tx Conclusão), principalmente no que se refere a financiamento (crédito estudantil), a sexo, a origem (escola pública/privada) e a utilização da nota do Enem na seleção para ingresso, apresenta aumento em sua importância em comparação com a variável taxa de concorrência (um modelo mais equilibrado). Aqui, semelhantemente ao que ocorreu em relação às IES públicas, nota-se ainda mais, como diferença, um crescimento na influência de fatores como faixa etária e motivação de escolha da IES, em contraposição ao turno e à duração do curso (que só se mostrou mais importante no modelo do XGB) que caíram no nível de influência (talvez por estarem mais ligados à evasão do que à retenção).

Outrossim, verifica-se novamente para as IES privadas que fatores institucionais passam a ter mais relevância (critérios de escolha das IES), enquanto que semelhantemente às IES públicas também são importantes fatores que refletem o prestígio social do curso (concorrência, processo seletivo pelo Enem) e as condições pessoais do aluno (sexo, faixa etária e formação do ensino básico). A integração acadêmica (atividades extracurriculares) continua também sendo um fator de marcante influência sobre o desempenho dos cursos desse segmento.

CONCLUSÃO

O objetivo deste estudo foi identificar e analisar, por meio de técnicas de *Machine Learning*, os fatores que influenciam no desempenho de um curso de graduação, especialmente quanto às taxas de ocupação e conclusão. Visou-se

trazer contribuições e explorar o potencial de ML para avaliação de políticas públicas voltadas para a expansão do Ensino Superior no Brasil.

Os dados foram obtidos principalmente do censo da educação superior (INEP, 2022), tendo sido escolhido o ano de 2019 para as análises, a fim de evitar potenciais distorções devido ao período de pandemia pela covid-19.

A base do censo superior mostrou-se bastante rica em informações, assim pode-se endereçar alguns dos aspectos elencados em outros estudos, de cunho mais qualitativos (SALES JÚNIOR, 2013), ainda que indiretamente.

Obteve-se um nível de desempenho dos modelos (acurácia, precisão, *recall*, f1 score) entre 59% e 79%, o que indica um potencial promissor de tais modelos para a investigação do fenômeno sob estudo. Os resultados para os *targets* escolhidos são consistentes entre si (taxa de ocupação e taxa de conclusão), o que reforça a importância dos fatores identificados para o desempenho dos cursos.

A análise da importância dos fatores para o resultado preditivo dos modelos é outro aspecto que corrobora o potencial de assertividade de modelos baseados nas variáveis propostas. A taxa de concorrência é sempre identificada como fator de maior influência, o que era de se esperar, cursos com maior interesse e prestígio social trazem motivação maior para sua conclusão, além de que possuem ingresso de um aluno com melhor nível de formação no ensino básico. Destacam-se ainda fatores como rede de origem do aluno (escola pública ou privada), sexo e faixa etária de ingresso. A aplicação dos modelos mostrou ainda que alguns fatores de influência são bem distintos entre IES públicas e privadas, o que só enfatiza a necessidade de análise em separado para as políticas educacionais de ES em cada uma dessas redes.

Como sugestões de trabalhos futuros, pode-se testar outros algoritmos de ML com objetivo de obter modelos com grau maior de acurácia e, além disso, fazer análises longitudinais, avaliando os mesmos cursos ao longo de alguns anos, o que pode também aumentar o desempenho dos modelos e, conseqüentemente, a acurácia dos resultados. Também podem ser criados

modelos voltados à análise de desempenho de cursos na modalidade EaD e de grau de tecnólogo que não foram contemplados neste trabalho.

REFERÊNCIAS

BRASIL. **Lei n. 9.394, de 20 de dezembro de 1996.** Lei Darcy Ribeiro. Lei de Diretrizes e Bases da Educação Nacional - LDB. Estabelece as diretrizes e bases da educação nacional. Brasília, DF: Presidência da República, [2022]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l9394.htm. Acesso em: 20 ago. 2022.

BRASIL. **Lei n. 13.005, de 25 de junho de 2014.** Aprova o Plano Nacional de Educação e dá outras providências. Brasília, DF: Presidência da República, [2022]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l13005.htm. Acesso em: 20 ago. 2022.

FERNANDES, F. T. **Machine learning em saúde e segurança do trabalhador:** perspectivas, desafios e aplicações. 2021. Tese (Doutorado em Saúde Pública) – Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, 2021. Disponível em: https://www.teses.usp.br/teses/disponiveis/6/6143/tde-27012022-140548/publico/FernandesFT_DR_R.pdf. Acesso em: 20 ago. 2022.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning:** Data Mining, Inference and Prediction, 2. ed., Stanford: Springer, 2008. E-book. Disponível em: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.158.8831&rep=rep1&type=pdf>. Acesso em: 20 ago. 2022.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Regiões Geográficas do Brasil - 2017.** Rio de Janeiro: IBGE. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional/15778-divisoes-regionais-do-brasil.html?=&t=sobre>. Acesso em: 21 jul. 2022.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílio – 2019.** Rio de Janeiro: IBGE, 2019. Disponível em:

<https://www.ibge.gov.br/estatisticas/sociais/trabalho/17270-pnad-continua.html?edicao=28203&t=resultados>. Acesso em: 5 ago. 2022.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Estimativas de população enviadas ao TCU - 2021**. Rio de Janeiro: IBGE, 2021. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=31451&t=resultados>. Acesso em: 18 jul. 2022.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do Censo da Educação Superior – 2009 – 2020, atualizado em abril de 2022**. Brasília: INEP, 2022. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>. Acesso em: 3 jun. 2022.

PANDALA, S. R. **Lazy Predict Docs**. 2020. Disponível em: <https://lazypredict.readthedocs.io/en/latest/readme.html>. Acesso em: 10 set. 2022.

QIFANG BI. *et al.* What is Machine Learning? A Primer for the Epidemiologist, **American Journal of Epidemiology**, v. 188, n. 12, p. 2222–2239, dez. 2019. Disponível em: <https://doi.org/10.1093/aje/kwz189>. Acesso em: 20 ago. 2022.

SALES JÚNIOR, J. S. **Uma análise estatística dos fatores de evasão e permanência de estudantes de graduação presencial da UFES**. 2013. Dissertação (Mestrado em Gestão Pública) – Centro de Ciências Jurídicas e Econômicas, Universidade Federal do Espírito Santo (UFES), Vitória/ES, 2013. Disponível em: <http://repositorio.ufes.br/bitstream/10/1935/1/Dissertacao%20Jaime%20Sales%20Junior.pdf>. Acesso em: 29 mai. 2022.



TEMA: AUDITORIA

FERRAMENTA:
APRENDIZADO DE MÁQUINA NÃO
SUPERVISIONADO

ÍNDICE DE PRIORIZAÇÃO DE OBJETOS DE AUDITORIA: UM ESTUDO DE CASO PARA MUNICÍPIOS SERGIPANOS¹

JOSÉ LEONARDO RIBEIRO NASCIMENTO⁽⁸⁾

INTRODUÇÃO

A Controladoria-Geral da União (CGU) é o órgão de controle interno do governo federal responsável por realizar atividades relacionadas à defesa do patrimônio público e ao incremento da transparência da gestão, por meio de ações de auditoria pública, correição, prevenção e combate à corrupção e ouvidoria.

A Instrução Normativa n. 3, de 9 de junho de 2017, que aprova o Referencial Técnico da Atividade de Auditoria Interna Governamental do Poder Executivo Federal, estabelece que as unidades de auditoria interna devem “estabelecer um plano baseado em riscos para determinar as prioridades da auditoria, de

1 Este artigo resulta de trabalho de conclusão orientado pelo Professor Aléssio Tony C. Almeida.

forma consistente com objetivos e metas institucionais da Unidade Auditada”². Seguindo a mesma linha, o Manual de Orientações Técnicas da Atividade de Auditoria Interna Governamental do Poder Executivo Federal (MOTE), aprovado pela Instrução Normativa SFC n. 8, de 6 de dezembro de 2017, é enfático ao falar da centralidade do planejamento baseado em riscos para a atividade de auditoria interna governamental, ressaltando a necessidade de que ele seja “capaz de garantir que as Unidades de Auditoria Interna Governamental (UAIG) concentrem seus esforços nas áreas e atividades cujo alcance dos objetivos pode ser mais fortemente impactado por eventos internos ou externos, ou seja, nas áreas de maior risco”³.

Cada uma das Unidades Regionais da CGU elabora anualmente o seu Plano de Auditoria Interna (Paint), também chamado de Plano Operacional, em que devem ser contempladas todas as atividades de competência da Unidade Regional a serem realizadas naquele ano, incluindo as auditorias que devem ser planejadas, conforme já exposto, levando-se em consideração os riscos envolvidos. O Universo de Auditoria das Unidades Regionais da CGU é bastante vasto, incluindo diferentes órgãos e entidades federais, tais como os Institutos Federais de Ensino Superior, os Conselhos Regionais de classe, as unidades do sistema “S”, além dos entes estaduais e municipais — governo estadual, prefeituras e secretarias e fundos estaduais e municipais que recebem recursos federais para diferentes áreas, como saúde, educação, assistência social, urbanismo etc. Somam-se ainda a essa lista as transferências voluntárias, normalmente operacionalizadas mediante convênio, contrato de repasse ou termos de compromisso.

A tarefa de fazer um planejamento de auditoria baseado em risco nas Unidades Regionais da CGU é, portanto, bastante complexa, já que cada trabalho possui suas particularidades, em virtude do ente receptor do recurso, da forma de financiamento e das regras próprias de cada programa/ação governamental. Atualmente, o planejamento de auditoria nas Unidades Regionais

2 Instrução Normativa n. 3/2017, Seção II, Parágrafo 83.

3 Manual de Orientações Técnicas da Atividade de Auditoria Interna Governamental do Poder Executivo Federal, p. 10.

da CGU é realizado a partir de diversas abordagens: para algumas unidades federais é adotado o mapeamento dos macroprocessos; para outras, é levada em consideração uma avaliação mais subjetiva dos auditores, que leva em conta a experiência oriunda de auditorias anteriores. Para fiscalizar os municípios, contudo, não há um sistema objetivo definido e em uso que leve em conta os fatores de risco inerentes a cada ação governamental, de sorte que o planejamento acaba sendo, em boa parte, reativo, ou seja, insere-se esta ou aquela ação governamental no planejamento em função de uma demanda externa, uma notícia alarmante na mídia ou uma transferência mais relevante de recurso para determinado município.

Quando não há avaliação de riscos realizada pela Unidade Auditada nem for possível que a própria CGU realize essa avaliação, o Manual de Orientações Técnicas da Atividade de Auditoria Interna Governamental do Poder Executivo Federal apresenta uma alternativa para a seleção dos trabalhos de auditoria com base em riscos: associar os objetos de auditoria a **fatores de riscos**, que nada mais são que os critérios de priorização definidos pela CGU, de acordo com o seu objeto de auditoria. Essa seleção acontece em dois estágios: **definição dos fatores de risco e priorização dos objetos de auditoria**. Como já foi explicado, considerando a complexidade do trabalho da CGU, priorizar os objetos de auditoria não é tarefa simples.

A proposta do presente trabalho foi a criação do **Índice de Priorização de Objetos de Auditoria (IPOA)**, obtido a partir de um modelo matemático-estatístico que agrupa e prioriza municípios, estados, programas ou ações governamentais com base em fatores de risco previamente determinados, para auxiliar as Unidades Regionais da CGU no planejamento do seu Plano de Auditoria Interna. O IPOA é elaborado a partir da análise de dados de uma série de indicadores, que funcionam como fatores de risco para as áreas que se deseja auditar. O modelo foi desenvolvido de maneira que possa ser utilizado por outras Unidades Regionais da CGU, mas, para o presente trabalho, foi realizado um estudo de caso que se restringiu aos municípios do Estado de Sergipe. Além disso, o **IPOA** calculado neste trabalho contemplou **indicadores/fatores de risco relacionados às três principais áreas sociais:**

Saúde (Atenção Primária), Educação (Educação Básica) e Assistência Social, referentes ao período de **janeiro/2020 a dezembro/2020**, além de **indicadores/fatores de risco de vulnerabilidade à fraude**. A escolha das três áreas sociais decorre da sua relevância, tanto do ponto de vista orçamentário e financeiro quanto social. A título exemplificativo, no exercício de 2022, as funções **saúde, educação e assistência social** responderam por 82% do valor das transferências legais, voluntárias e específicas para o Estado de Sergipe e seus municípios, com ampla prevalência da área de **saúde**, que sozinha corresponde a 73,4% das transferências. Do ponto de vista social, essas três áreas incluem programas e ações de grande impacto que, por essa mesma razão, são costumeiramente avaliados pela CGU, tais como: Piso de Atenção Primária à Saúde, Promoção da Assistência Farmacêutica e Insumos Estratégicos na Atenção Básica em Saúde, Ações de Vigilância em Saúde, Apoio à Alimentação Escolar na Educação Básica (PNAE), Dinheiro Direto na Escola para a Educação Básica, Apoio ao Transporte Escolar na Educação Básica, Auxílio Brasil, Ações de Proteção Social Básica e Ações de Proteção Social Especial, dentre outros.

Foram utilizados **33 indicadores**, representando fatores de risco, consolidados em quatro áreas: **Vulnerabilidade à Fraude**, composto por quinze indicadores; **Vulnerabilidade - Saúde**, com cinco indicadores; **Vulnerabilidade - Educação**, com sete indicadores, e **Vulnerabilidade - Socioeconômica**, com seis indicadores. O **IPOA** foi gerado para cada município sergipano a partir da junção das quatro áreas. De posse do **IPOA**, as Unidades Regionais da CGU terão informações criteriosas e de qualidade para planejar parte relevante do seu trabalho anual de auditoria — avaliação dos programas e ações federais executadas no âmbito municipal —, em observância àquilo que preceituam os normativos vigentes e as boas práticas de auditoria.

Ressalta-se o alto nível de versatilidade do modelo, que pode ser adaptado de acordo com a necessidade de cada Unidade Regional da CGU, sendo possível, por exemplo, gerar o IPOA para uma área ou ação específica (IPOA Educação Básica ou IPOA Urbanismo, por exemplo), substituir fatores de risco de acordo com a disponibilidade dos dados, adicionar áreas, ou mesmo, testar

novas maneiras de obter o IPOA. A necessidade de cada Unidade Regional determinará qual a melhor forma de utilizar o índice.

AUDITORIA BASEADA EM RISCOS

A auditoria interna baseada em riscos une a auditoria interna ao processo de gerenciamento de riscos, “permitindo que o foco dos trabalhos realizados esteja voltado para os principais riscos de uma organização não atingir os objetivos para os quais existe” (OLIVEIRA, 2019). Dentro do contexto das organizações, o risco está sempre associado aos objetivos. A ISO 31.000/2018 define o risco como o “efeito da incerteza nos objetivos”, enquanto o *Committee of Sponsoring Organizations of the Treadway Commission* (COSO) diz que risco é “a possibilidade de que um evento ocorrerá e afetará negativamente a realização dos objetivos”⁴.

Considerando a escassez dos recursos e a amplitude do universo de auditoria da CGU, conforme já demonstrado, é fundamental que os trabalhos de auditoria sejam planejados com base em riscos, para que sejam priorizadas justamente aquelas ações mais relevantes, que apresentem maior risco de comprometer os objetivos da CGU e das unidades auditadas. O MOTE apresenta três diferentes bases que podem ser utilizadas para planejamento da auditoria com base em riscos, a serem utilizadas prioritariamente, conforme ordem a seguir:

1. Avaliação de riscos realizada pela Unidade Auditada.
2. Avaliação de riscos realizada pela UAIG.
3. Fatores de risco.

Na CGU, o planejamento de auditoria para os programas e as ações federais executados por entes municipais leva em consideração os **fatores de risco**, ou seja, os critérios de priorização aplicáveis segundo cada ação a ser auditada.

4 COSO, 1992.

FATORES DE RISCO

A primeira parte do trabalho consistiu na escolha dos indicadores a serem utilizados, representando fatores de risco referentes a cada área — **saúde, educação e assistência social** —, além de indicadores de **vulnerabilidade à fraude**. Quanto mais alto o valor do indicador, maior o risco de que o objetivo dos programas e das ações referentes àquela área não sejam alcançados. Optou-se, neste estudo de caso, pela criação do **Índice de Priorização de Objetos de Auditoria** amplo, que contempla uma avaliação acerca das três principais áreas sociais dos municípios, mas isso pode ser mudado de acordo com a necessidade de cada Unidade Regional: se a orientação é realizar auditorias preferencialmente na área de saúde, por exemplo, basta não utilizar os indicadores de educação, calculando-se o **“IPOA Saúde”**, utilizando-se apenas os indicadores de vulnerabilidade na saúde, vulnerabilidade socioeconômica e vulnerabilidade à fraude.

Vulnerabilidade à Fraude

Os indicadores de vulnerabilidade à fraude foram escolhidos a partir dos testes e das avaliações que são realizados costumeiramente nas auditorias da CGU, seja em programas federais, seja em convênios ou unidades jurisdicionadas. Relacionam-se a fatores de risco ligados à **gestão municipal**, aos **servidores municipais**, especialmente àqueles em cargo de comissão, à **gestão de compras**, especialmente licitações, dispensas e inexigibilidades de licitação, e, finalmente, aos fatores de risco ligados aos **contratados** para fornecer bens e serviços para a administração pública municipal. O quadro a seguir detalha os fatores de risco de vulnerabilidade à fraude utilizados:

Quadro 1 | Fatores de Risco — Vulnerabilidade à Fraude

Indicador	Indicador
<p>F1 – Proporção de ordenadores de despesas que têm participação societária em empresas.</p> <p>F2 – Proporção de servidores da prefeitura que têm participação societária em empresas.</p> <p>F3 – Proporção de Ordenadores de Despesa que estavam filiados a partidos políticos em 2021.</p> <p>F4 – Proporção de beneficiários de empenho filiados a partidos políticos na última eleição para prefeito.</p> <p>F5 – Proporção de Ordenadores de Despesa que fizeram doação a partidos políticos na última eleição para prefeito.</p> <p>F6 – Proporção de beneficiários de empenho que fizeram doações a partidos políticos na última eleição para prefeito.</p> <p>F7 – Proporção de empresas beneficiárias de empenho inseridas no Cadastro de Empresas Inidôneas e Suspensas no exercício de 2021.</p> <p>F8 – Proporção de empresas beneficiárias de empenho relacionadas no Cadastro de Entidades Privadas sem Fins Lucrativos Impedidas.</p>	<p>F9 – Proporção de beneficiários de empenho com CPF cancelado.</p> <p>F10 – Proporção de beneficiários de empenho com CNPJ baixado.</p> <p>F11 – Proporção de empresas beneficiárias de empenho com menos de três empregados.</p> <p>F12 – Proporção de empresas beneficiárias de empenho com sócios com remuneração na GFIP inferior a R\$ 1.100,00 entre janeiro/2018 e Dezembro/2021.</p> <p>F13 – Proporção de empresas cujo capital social é inferior a 10% do valor total empenhado no município.</p> <p>F14 – Proporção de dispensas e inexigibilidades em relação à soma das licitações, dispensas e inexigibilidades do município.</p> <p>F15 – Proporção do valor total de pregões presenciais realizados pelo município em relação à soma de pregões presenciais e pregões eletrônicos.</p>

Fonte: Elaborado pelo autor

Vulnerabilidade — Saúde

Os fatores de risco referentes à saúde limitaram-se, neste trabalho, à Atenção Primária e foram divididos em dois grupos: **oferta** e **resultado** dos serviços de saúde na atenção básica. Optou-se por utilizar indicadores clássicos dessas duas áreas, conforme detalhado no quadro a seguir:

Quadro 2 | Fatores de Risco — Vulnerabilidade — Saúde

Indicador (Oferta de serviços de Saúde)	Indicador (Resultado de serviços de Saúde)
S1 – Proporção de médicos em especialidades da atenção primária por 1000 habitantes. S2 – Cobertura de Equipes de Saúde da Família.	S3 – Mortalidade Infantil. S4 – Mortalidade por causas evitáveis na população de 5 a 74 anos. S5 – Proporção de Internações por Condições Sensíveis à Atenção Primária.

Fonte: Elaborado pelo autor

Vulnerabilidade — Educação

Os fatores de risco referentes à Educação limitaram-se à Educação Básica e foram selecionados a partir da relação de indicadores educacionais disponibilizados em dados abertos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), conforme detalhado no quadro a seguir:

Quadro 3 | Fatores de Risco — Vulnerabilidade — Educação

Indicador	Observação
E1 – Nível do Indicador de complexidade de gestão da escola. E2 – Média de Alunos por Turma do Ensino Fundamental. E3 – Número médio de Horas-Aula Diária no Ensino Fundamental. E4 – Média do Indicador de Nível Socioeconômico (Inse) de 2019.	E5 – Taxa de Distorção Idade-Série do Ensino Fundamental. E6 – Taxas de Rendimento Escolar – Reprovação do Ensino Fundamental. E7 – Taxas de Transição – Evasão do Ensino Fundamental.

Fonte: Elaborado pelo autor

Vulnerabilidade socioeconômica

Os fatores de risco socioeconômicos são mais amplos, podendo ser utilizados como referência para a área de Assistência Social, bem como para

indicar a vulnerabilidade socioeconômica dos municípios. Os indicadores estão detalhados no quadro a seguir:

Quadro 4 | Fatores de Risco — Vulnerabilidade socioeconômica

Indicador	Observação
SE1 – Grau de instrução dos eleitores do Brasil. SE2 – Proporção de famílias em situação de pobreza inscritas no CadÚnico. SE3 – Proporção de famílias em situação de extrema pobreza inscritas no CadÚnico.	SE4 – Proporção de idosos (65 anos ou mais) sobre população. SE5 – Razão de dependência. SE6 – Arrecadação total.

Fonte: Elaborado pelo autor.

COLETA E TRATAMENTO DOS DADOS

Os dados foram obtidos de diversas fontes, tais como Datasus – CNES, e-Gestor AB, CGUData, Inep, TSE, Siconfi e CadÚnico. Optou-se por utilizar o ano de 2020 como referência, tendo em vista a indisponibilidade de informações referentes ao ano de 2021 para alguns dos indicadores utilizados, tais como aqueles do grupo **Vulnerabilidade — Fraude**. Para esse grupo, aliás, ressalta-se que os dados não estão disponíveis publicamente, apenas na base de dados **CGUData**, de uso exclusivo dos servidores da CGU. Dessa forma, para a divulgação do **IPOA**, atribuiu-se um código aleatório para cada município, de forma a impossibilitar a sua identificação.

Os fatores de risco foram consolidados nas quatro áreas de duas maneiras:

- 1) por **média simples**, após terem seus valores normalizados entre 0 e 1; e
- 2) por **média ponderada**, considerando pesquisa feita junto a auditores especialistas da CGU, que avaliaram a relevância de cada indicador, atribuindo pesos entre 1 e 5.

Tendo em vista a evidente singularidade de Aracaju dentre os municípios sergipanos, a capital sergipana foi excluída do modelo, que, dessa forma, compreende os 74 municípios do interior.

MODELOS TESTADOS

Foram realizados diversos testes para ajustar o modelo e avaliar a sua robustez, de forma a obter o **Índice de Priorização dos Objetos de Auditoria** mais confiável, considerando os fatores de risco utilizados. Para tanto, foram testados sete modelos, ressaltando-se que para cada um deles foram realizados testes com a média simples (indicadores sem peso) e média ponderada (indicadores com peso), resultando, portanto, em 14 modelos distintos. Os modelos utilizados foram os seguintes:

Modelo 1: Clusterização pela Vulnerabilidade socioeconômica com atribuição de nota a partir da média dos outros índices — Após realizar a clusterização pela **Vulnerabilidade socioeconômica** em 8 *clusters*, o **IPOA** foi calculado para cada um dos municípios por meio da soma dos outros índices, conforme fórmula a seguir:

$$\mathbf{IPOA} = \mathit{Vuln. \grave{a} Fraude} + \mathit{Vuln. na Sa\acute{u}de} + \mathit{Vuln. na Educa\c{c}\~{a}o}$$

Modelo 2: Clusterização pela Vulnerabilidade socioeconômica, com atribuição de nota a partir da média das outras áreas e com $k = 4$ – Similar ao modelo 1, mas com 4 *clusters*.

Modelo 3: Sem clusterização, com soma simples das quatro áreas — O **IPOA** foi calculado pela soma simples dos índices das quatro áreas, conforme fórmula a seguir:

$$\mathbf{IPOA} = \mathit{Vuln. \grave{a} Fraude} + \mathit{Vuln. na Sa\acute{u}de} + \mathit{Vuln. na Educa\c{c}\~{a}o} + \mathit{Vuln. Socioecon\~{o}mica}$$

Modelo 4: Sem clusterização, com cálculo do IPOA a partir da soma dos índices das áreas de fraude, saúde e educação e ponderação pelo valor do índice socioeconômico:

$$IPOA = (Vuln. \text{ à Fraude} + Vuln. \text{ na Saúde} + Vuln. \text{ na Educação}) * Vuln. \text{ Socioeconômica}$$

Modelo 5: Sem clusterização, com obtenção do IPOA a partir da soma dos índices das áreas saúde, educação e socioeconômica e ponderação pelo valor do índice de vulnerabilidade à fraude:

$$IPOA = (Vuln. \text{ Socioeconômica} + Vuln. \text{ na Saúde} + Vuln. \text{ na Educação}) * Vuln. \text{ à Fraude}$$

Modelo 6: Clusterização pela Vulnerabilidade à fraude, com atribuição de nota a partir da média dos outros índices:

$$IPOA = Vuln. \text{ na Saúde} + Vuln. \text{ na Educação} + Vuln. \text{ Socioeconômica}$$

Modelo 7: Clusterização pela Vulnerabilidade à fraude, com atribuição de nota a partir da média dos outros índices e $k = 4$:

$$IPOA = Vuln. \text{ na Saúde} + Vuln. \text{ na Educação} + Vuln. \text{ Socioeconômica}$$

RESULTADOS DOS MODELOS

Verificou-se que há consistência entre os resultados dos diferentes modelos. Apenas 26 dos 74 municípios de Sergipe que foram incluídos na análise apareceram nas 14 listas de 10 (dez) municípios. Cinco municípios apareceram em dez ou mais modelos. A análise dos dados aponta ainda uma limitação

dos modelos 1 e 2, baseados na clusterização pela **Vulnerabilidade socioeconômica**: dois municípios com IPOA bastante baixos — 2,6 e 3,5 — foram selecionados porque apareciam como únicos integrantes de seus *clusters*, tendo em vista as suas características socioeconômicas únicas. Os modelos 6 e 7, que também basearam-se em clusterização, mas a partir da **Vulnerabilidade à fraude**, superaram essa limitação, já que não houve municípios com IPOA baixos. Outra observação é a tendência de que os municípios com os maiores IPOA médios tenham sido selecionados por vários modelos. A única exceção entre os dez municípios com o maior IPOA médio é o município 645378, que figura em 5º lugar, com um IPOA de 8,10, mas que foi selecionado apenas no modelo 7.

Mas, afinal, qual o melhor modelo? Uma vez que foram muitos os modelos testados, optou-se por selecionar aquele que apresenta resultados mais robustos, selecionando os municípios com maior prioridade de auditoria nos diferentes modelos testados. Destarte, o primeiro passo foi descartar os modelos que usaram os indicadores sem o peso estabelecido pelos auditores especialistas da CGU, o que resultou em sete modelos. Em seguida, foram utilizados dois critérios para escolher o modelo mais adequado, todos baseados no número de municípios com maior **IPOA** selecionados por cada modelo:

SOMA DO IPOA

A partir da lista agregada dos municípios, obtida com a aplicação dos sete diferentes modelos, foram selecionados os dez municípios com a maior soma do IPOA. Dessa forma, considera-se não apenas a frequência com que determinado município é selecionado pelos modelos, como também o seu Índice de Priorização de Objetos de Auditoria. Os municípios selecionados foram os seguintes:

Tabela 1 | Dez municípios com maior soma do IPOA

Município	Soma do IPOA
925739	64,33
565618	64,28
871352	61,40
513159	55,67
896083	49,36
191608	39,22
313942	34,53
981261	32,08
909630	26,61
821167	18,78

Fonte: Elaborado pelo autor.

O passo seguinte foi verificar qual modelo selecionou o maior número de municípios presentes na tabela 2. O resultado foi o seguinte:

Tabela 2 | Municípios selecionados por cada modelo

Modelo	Municípios
3	9
4	8
5	8
6	8
1	7
7	7
2	6

Fonte: Elaborado pelo autor.

Há bastante equilíbrio, como se pode notar, com três modelos selecionando oito dos dez municípios e o **Modelo 3** (*Sem clusterização, com soma simples das quatro áreas*) selecionando nove dos dez municípios da lista.

MÉDIA DO IPOA

O outro critério foi selecionar os dez municípios considerando a média do IPOA, obtida dos diferentes modelos em que eles foram selecionados. Os dez municípios com maior média foram os seguintes:

Tabela 3 | Dez municípios com maior média do IPOA

Município	IPOA médio
925739	9,19
565618	9,18
871352	8,77
896083	8,23
645378	8,10
981261	8,02
513159	7,95
191608	7,84
612684	7,02
313942	6,91

Fonte: Elaborado pelo autor.

Ressalta-se que oito municípios aparecem nas duas listas (tabelas 5 e 6). O passo seguinte foi verificar qual modelo selecionou o maior número de municípios presentes nessa lista, com o seguinte resultado:

Tabela 4 | Municípios selecionados por cada modelo

Modelo	Municípios
3	8
7	8
1	7
4	7
5	7
6	7
2	6

Fonte: Elaborado pelo autor.

Novamente, o **Modelo 3** (*Sem clusterização, com soma simples das quatro áreas*), destacou-se, dessa vez ao lado do **Modelo 7** (*Clusterização pelo IRA Fraude, com atribuição de nota a partir da média dos outros índices e $k = 4$*), ambos selecionando oito dos dez municípios da lista.

A partir desses dois critérios, portanto, o **Modelo 3 — Sem clusterização, com soma simples das quatro áreas e com utilização de indicadores com pesos atribuídos pelos especialistas da CGU** — mostrou-se como o mais robusto e consistente, resultando na seleção de municípios com maior prioridade de auditoria nas áreas de saúde, educação e assistência social e com maior vulnerabilidade à fraude.

O **Modelo 3** apresentou IPOA médio de 8,33 para os dez municípios, a maior média dentre os sete modelos que usaram os indicadores com peso. Todos os municípios ocupam o primeiro lugar em relação a pelo menos um dos 33 indicadores. O município **925739**, com maior IPOA do modelo, ocupa a 3ª colocação geral na área Vulnerabilidade à fraude e na área Vulnerabilidade na educação e a 9ª colocação na Vulnerabilidade socioeconômica. O município **565618**, por sua vez, ocupa o primeiro lugar em relação aos indicadores *F3 — Proporção de Ordenadores de Despesa que estavam filiados a partidos políticos em 2021*, *F5 — Proporção de Ordenadores de Despesa que fizeram doação a partidos políticos na última eleição para prefeito*, *F15 — Proporção do valor total de pregões presenciais realizados pelo município em relação à*

soma de pregões presenciais e pregões eletrônicos e E6 — Taxas de Rendimento Escolar — Reprovação do Ensino Fundamental. Sete municípios estão entre os dez com maior Vulnerabilidade à fraude, incluindo os três primeiros colocados. Além disso, o modelo capturou o município com maior vulnerabilidade na saúde (**191608**) e na educação (**565618**), além do município com a terceira maior vulnerabilidade social (**896083**). Restou, portanto, evidente, que os dez municípios selecionados pelo modelo 3 apresentam indicadores que os colocam como as maiores prioridades de auditoria para a CGU-Regional dentre os 74 municípios analisados.

CONCLUSÃO

Realizar auditorias baseadas em risco é um desafio para as Unidades Regionais da CGU, haja vista o amplo espectro de ações sob sua competência. No caso da fiscalização dos recursos federais transferidos para os municípios, esse problema torna-se ainda mais difícil, visto que são muitos municípios, cada um com características distintas, diversos programas e ações, sendo uma tarefa bastante complexa selecionar e avaliar fatores de risco que levem em conta riscos como fraude, ineficiência na oferta de serviços, deficiência nos resultados da ação e riscos socioeconômicos próprios de cada município.

O **Índice de Priorização de Objetos de Auditoria** ora apresentado utiliza-se da ciência de dados para se tornar uma ferramenta bastante útil para a Controladoria-Geral da União, haja vista apresentar uma opção de planejamento anual para as Unidades Regionais que leva em conta fatores de risco consistentes para as mais diversas áreas, como, no presente estudo de caso, saúde, educação e assistência social, que comumente são objeto de auditoria da CGU. Além disso, como já explicado, é possível utilizar o IPOA de maneira ampla, bem como adaptar seu modelo para outras áreas, sendo necessário apenas escolher os indicadores/fatores de risco adequados. Reforça-se que apesar de o estudo de caso contemplar apenas as transferências destinadas aos municípios de Sergipe, o modelo foi pensado para poder ser facilmente

adaptado para outras Unidades Regionais, como já longamente demonstrado no decorrer do trabalho.

A utilização do IPOA possibilita que a CGU planeje seus trabalhos em consonância com as melhores práticas de auditoria atualmente vigentes, levando-se em conta fatores de risco, ao mesmo tempo que, quando se utilizar da ciência de dados, permite que o trabalho seja feito em larga escala, contemplando um número elevado de indicadores e, conseqüentemente, de áreas analisadas — no presente trabalho, por exemplo, foram utilizados 33 indicadores de diversas áreas.

Ressalta-se que não é intenção do modelo apresentar uma lista fechada e exaustiva dos municípios, mas subsidiar o planejamento da Unidade Regional com informações robustas e de qualidade, que deem sustentação metodológica ao seu Plano Anual de Auditoria Interna.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 31000: Gestão de riscos – princípios e diretrizes**. Rio de Janeiro. 2009.

BRASIL. Conselho de Monitoramento e Avaliação de Políticas Públicas. **Relatório de Avaliação – Atenção Primária à Saúde – APS (Ciclo 2020)**. Brasília, 2020.

_____. Ministério da Saúde. **Portaria n. 2.436, de 21/09/2017**. Política Nacional de Atenção Básica. Brasília, 2017.

_____. Ministério do Planejamento, Orçamento e Gestão (MP); Ministério da Transparência e Controladoria-Geral da União. **Instrução Normativa Conjunta n. 1, de 10 de maio de 2016**. Brasília: Diário Oficial da União, 11 maio 2016.

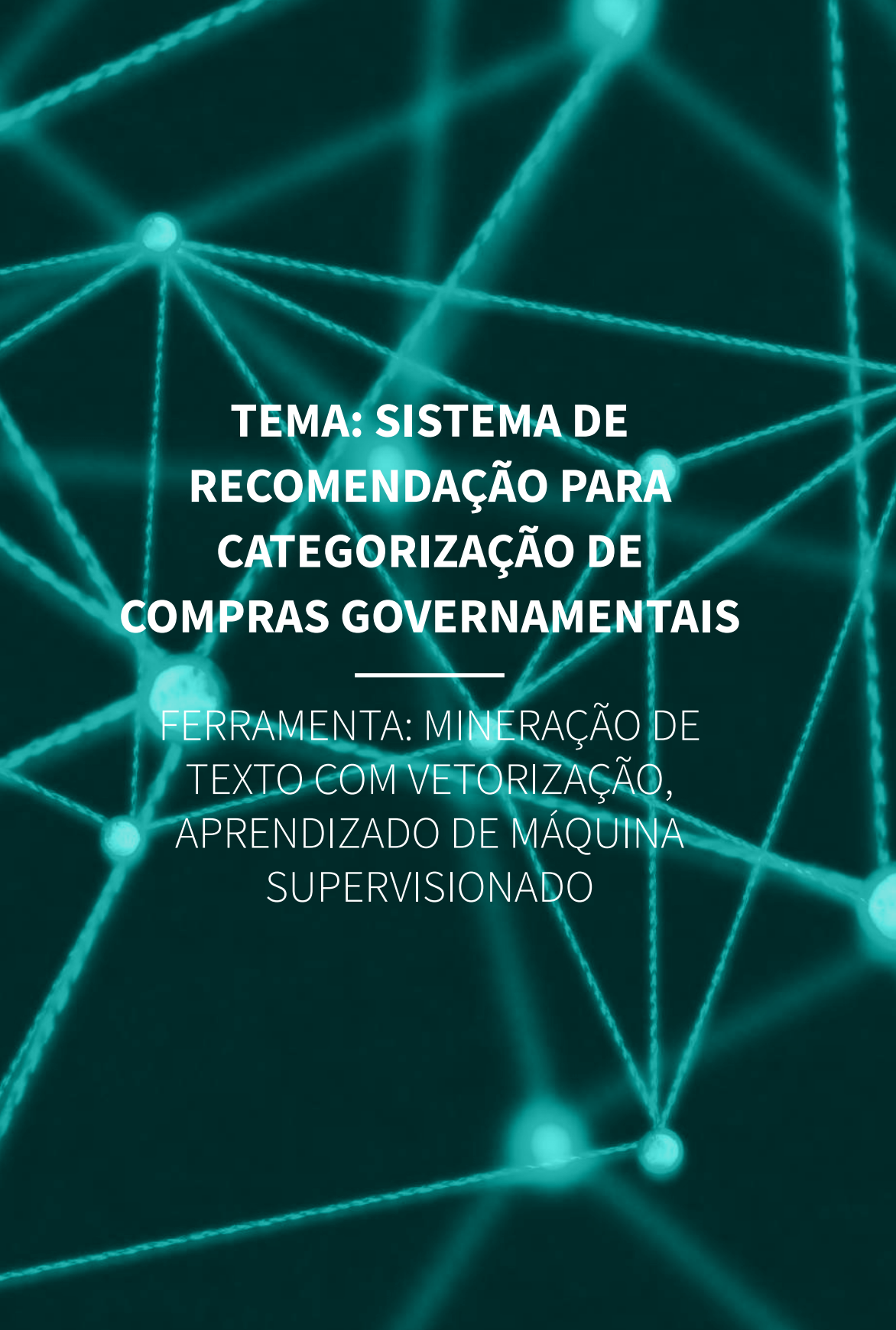
_____. Secretaria Federal de Controle Interno. **Instrução Normativa n. 3, de 9 de junho de 2017**. Referencial Técnico da Atividade de Auditoria Interna Governamental do Poder Executivo Federal. Brasília, 2017.

_____. Secretaria Federal de Controle Interno. **Instrução Normativa n. 8, de 8 de dezembro de 2017**. Manual de Orientações Técnicas da Atividade de Auditoria Interna Governamental do Poder Executivo Federal. Brasília, 2017.

_____. Secretaria Federal de Controle Interno. **Orientação Prática: Plano de auditoria interna baseado em riscos**. Brasília, 2020.

_____. Tribunal de Contas da União. **Instrução Normativa n. 63, de 1 de setembro de 2010**. Brasília, 2010.

OLIVEIRA, T. F. de. Avaliação das práticas de auditoria interna da Secretaria Federal de Controle Interno da CGU sob a ótica da auditoria baseada em riscos. **REVISTA DA CGU**, Brasília, v. 11 n. 19 (2019), 19, abr-jul/2019. Disponível em: <https://repositorio.cgu.gov.br/handle/1/44424>. Acesso em: jul. 2022.



**TEMA: SISTEMA DE
RECOMENDAÇÃO PARA
CATEGORIZAÇÃO DE
COMPRAS GOVERNAMENTAIS**

FERRAMENTA: MINERAÇÃO DE
TEXTO COM VETORIZAÇÃO,
APRENDIZADO DE MÁQUINA
SUPERVISIONADO

CLASSIFICAÇÃO AUTOMÁTICA DE COMPRAS LICITADAS PELO SISTEMA COMPRAS.GOV.BR¹

MATHEUS SCATOLINO DE REZENDE⁽⁹⁾

INTRODUÇÃO

As compras governamentais são aquisições de bens e serviços, realizadas pela Administração Pública, de todas as esferas e poderes a partir da oferta de fornecedores privados ou públicos (ex.: empresas estatais). Destaca-se que as compras governamentais no Brasil representam elevada fatia do PIB, estimada em 12,5% no período de 2006 até 2016 (RIBEIRO; INÁCIO JÚNIOR, 2019).

O uso da tecnologia da informação tem um papel preponderante na racionalização e na desburocratização das compras da Administração Pública. Nesse contexto, o sistema Compras.gov.br (antigo Comprasnet) foi desenvolvido para suportar as etapas da contratação pública: Planejamento da Contratação, Seleção do Fornecedor e Gestão e Fiscalização do Contrato (“Conheça o Compras.gov.br,” [s.d.]). Ressalta-se que esse sistema é de uso obrigatório pela Administração Direta do Poder Executivo Federal. Os estados e os municípios também são obrigados a utilizarem o Compras.gov.br quando executarem recursos da União decorrentes de transferências voluntárias.

1 Este artigo resulta de trabalho de conclusão orientado pelo Professor Alex Lopes Pereira.

O módulo Catálogo é uma base de dados que categoriza todos os bens e serviços licitados pelo sistema Compras.gov.br. O Catálogo é dividido por materiais (CATMAT) e serviços (CATSER). O pregoeiro deve definir uma categoria do CATMAT/CATSER para cada item a ser adquirido pelo governo no sistema Compras.gov.br. Entretanto, a atribuição dos itens em uma categoria se sujeita a erros humanos, considerando especialmente a grande quantidade de categorias (atualmente existem 141.626 categorias de materiais e 2.683 categorias de serviços).

A adequada identificação de compras públicas em uma categoria do CATMAT ou do CATSER é fundamental para a pesquisa de preços na Administração Pública. A categoria é utilizada para identificar compras passadas e estabelecer o valor estimado de determinado tipo de material ou serviço. Portanto, a melhoria da classificação das compras evita o desperdício de recursos públicos, considerando que a administração definirá um preço estimado com base em preços reais, efetivamente praticados junto à Administração Pública.

Por exemplo, a categoria “Acesso à Internet (Banda Larga)” deve conter todas as compras de serviços de acesso à internet da Administração Pública Federal. Caso essa categoria seja associada com compras de outros serviços (ex.: telefonia fixa), o gestor terá dificuldade em estimar o preço do serviço de acesso à internet. De forma semelhante, se um serviço de acesso à internet for associado a outra categoria, isso também diminuirá a precisão da estimativa de preços.

A estimativa de preços de compras públicas é utilizada para determinar o orçamento máximo estimado para a contratação. Conforme o Art. 59 da Lei n. 14.133/2021 (nova lei de licitações que substituiu a Lei n. 8.666/1993), as propostas dos fornecedores serão desclassificadas se permanecerem acima do orçamento estimado para a contratação. A forma preferencial para obter preços para definir o valor estimado da contratação é a mediana do item correspondente nos sistemas oficiais de governo (Art. 5º da IN n. 65/2021/SEGES/ME).

A definição de uma estimativa acima dos preços (superestimativa) praticados acarreta o risco de conluio entre os fornecedores. Eles podem se organizar para

não concorrerem entre si. Conseqüentemente, o valor superestimado da licitação não será reduzido, e a Administração Pública contratará um item ou serviço acima do preço de mercado, ocasionando desperdício de recursos públicos.

Além da perspectiva financeira, destaca-se a perspectiva do controle social, na qual a correta classificação das compras públicas permite uma visão mais geral dos gastos públicos a partir da agregação em categorias. Como exemplo, destacam-se os recorrentes questionamentos efetuados pela sociedade quanto à compra de produtos alimentícios considerados supérfluos pela Administração Pública.

Considerando que a classificação de compras públicas é um processo manual, o problema principal é a eventual associação incorreta entre uma categoria do Catálogo e o material ou o serviço que está sendo adquirido, considerando a ampla quantidade de categorias (atualmente existem 141.626 categorias de materiais e 2.683 categorias de serviços).

Além de incorrer em erros, a classificação manual é onerosa. O extenso catálogo demanda esforços substanciais dos pregoeiros para selecionar a categoria de cada item da licitação corretamente.

Portanto, essa pesquisa teve como objetivo desenvolver uma prova de conceito de um sistema de recomendação automatizado das categorias do Catálogo de Serviços (CATSER) de TI, a partir da descrição de novo item de compra com a utilização de um modelo de aprendizado de máquinas supervisionado multiclasse.

Essa prova de conceito deve permitir a implementação futura de interface em que o usuário informe a descrição de um item de compra e, a seguir, receba a sugestão das categorias mais prováveis, conforme figura 1.

Figura 1 | Protótipo da implementação futura deste estudo.

Informe a descrição do item da licitação para sugestões de categorias

Categorias sugeridas:

- Desenvolvimento de novo software - Java - 80% de probabilidade
- Desenvolvimento de novo software - Outras linguagens - 11% probabilidade
- Serviços de pesquisa, análise e desenvolvimento em tecnologia da informação e comunicação (TIC) - 4% probabilidade
- Serviços de consultoria em Tecnologia da Informação e Comunicação (TIC) - 3% de probabilidade
- Software como Serviço - SaaS - 2% de probabilidade

REFERENCIAL TEÓRICO

O aprendizado de máquina engloba algoritmos que combinam conhecimento de diversas áreas, como estatística, neurobiologia e inteligência artificial e permite que computadores simulem o aprendizado (NASTESKI, 2017). Esses algoritmos utilizam uma série de dados para detectar padrões que servem de estímulo para extrair informações de novos dados.

O aprendizado de máquina supervisionado abrange algoritmos que detectam padrões a partir de um treinamento com diversos exemplos de dados de entrada com os respectivos rótulos esperados (MAHESH, 2020). A partir desse treinamento, em um problema de classificação, infere-se uma função que pode ser utilizada para rotular novos dados, conforme figura 2. Atributos ou características (*features*) são parâmetros de cada amostra do conjunto de dados que podem contribuir para a elaboração de um modelo preditivo (SEN; HAJRA; GHOSH, 2020).

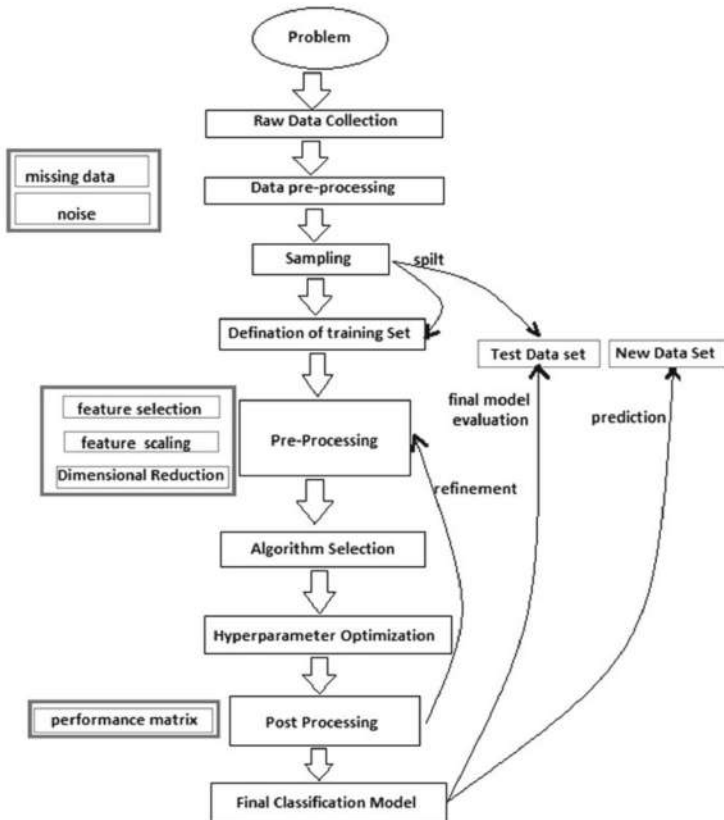
As compras públicas na Administração Pública Federal foram beneficiadas pela aplicação da tecnologia da informação. Nesse processo de transformação digital das compras públicas, destaca-se a criação do sítio Comprasnet em 1998. Assim, ampliou-se o acesso ao público e aos fornecedores pela internet e reduziram-se os custos de transação (FERNANDES, 2020).

O Compras.gov.br (antigo Comprasnet) é o principal portal de compras públicas do Brasil, considerando a obrigatoriedade de uso pela Administração Pública Federal direta, além de estados e municípios quando executam recursos da União decorrentes de transferências voluntárias, tais como con-

vênios e contratos de repasse, para a aquisição de bens e a contratação de serviços comuns, conforme Decreto n. 10.024/2019. Esse sistema atualmente oferece suporte a todas as etapas da contratação pública: Planejamento da Contratação, Seleção do Fornecedor e Gestão e Fiscalização do Contrato.

Cabe destacar que a qualidade dos dados relacionados com as compras públicas é um desafio significativo para os analistas de dados (SOYLU et al., 2022). Entre os principais potenciais problemas estão dados faltantes, errados, duplicados e pobremente formatados (SOYLU et al., 2022).

Figura 2 | Fluxograma de um modelo de classificação utilizando aprendizado de máquina supervisionado.



Fonte: SEN; HAJRA; GHOSH, 2020

METODOLOGIA

VISÃO GERAL

Esta pesquisa foi elaborada conforme a metodologia baseada no modelo CRISP-DM: realizar revisão sistemática da literatura relacionada com os temas de pesquisa, compreender o contexto do negócio e dos dados, realizar o pré-processamento dos dados e selecionar, treinar e comparar os resultados dos algoritmos supervisionados multiclasse.

O banco de dados utilizado para a extração dos dados nessa pesquisa foi oriundo do sistema DW-SIASG. Mediante a tecnologia *Data Warehouse*, esse sistema consolida e organiza as informações relativas às compras extraídas do Sistema Integrado de Administração de Serviços Gerais — SIASG.

Com o objetivo de medir o desempenho dos algoritmos, os dados de compras classificadas previamente e manualmente pelos pregoeiros foram divididos em duas partes: treinamento e testes. O primeiro será utilizado para treinar um modelo e o segundo será utilizado para medir o desempenho do modelo, considerando a premissa de que os falsos negativos são mais prejudiciais do que os falsos positivos nos resultados apresentados por esse modelo. Diferentes algoritmos serão avaliados e seus resultados comparados.

COMPREENSÃO DO NEGÓCIO

O Sistema de Contratações do Governo Federal — Compras.gov.br, evolução do Sistema Comprasnet —, é responsável pelo suporte de todas as etapas da contratação pública: Planejamento da Contratação, Seleção do Fornecedor e Gestão e Fiscalização do Contrato (“Conheça o Compras.gov.br,” [s.d.]). Esse sistema é de utilização obrigatória por todos os órgãos e as entidades do SISG.

Os estados e os municípios também são obrigados a utilizarem o Compras.gov.br quando executarem recursos da União decorrentes de transferências voluntárias, tais como convênios e contratos de repasse, para a aquisição de

bens, e a contratação de serviços comuns, conforme Decreto nº 10.024/2019. Destaca-se que esses entes podem utilizar facultativamente o Compras.gov.br para suportar as demais compras públicas.

Nesse contexto, o Catálogo de Materiais (CATMAT) e o Catálogo de Serviços (CATSER) são subsistemas do Compras.gov.br para padronizar os tipos de objetos de cada licitação e contratação dos órgãos do SISG. O CATMAT adota os critérios da metodologia *Federal Supply Classification*, utilizada pelo governo dos Estados Unidos, e o CATSER se baseia nos critérios adotados pela Organização das Nações Unidas (BRASIL, 2020).

COMPREENSÃO DOS DADOS

O CATMAT possui atualmente 141.626 categorias de materiais e o CATSER, 2.683 categorias de serviços. Neste estudo, optou-se por desenvolver uma prova de conceito para categorização automática dos serviços de Tecnologia da Informação (TI).

Cada categoria do CATSER possui um código hierárquico que agrupa os serviços conforme a metodologia *Central Product Classification* (CPC), adotada pela ONU, e adaptada pelo Compras.gov.br. A hierarquia desse código CPC é composta pelos níveis Seção, Divisão, Grupo, Classe e Subclasse de Serviço (BRASIL, 2020).

As licitações no Compras.gov.br são compostas por metadados a serem preenchidos obrigatoriamente pelo pregoeiro. Entre eles, destacam-se o código da unidade (UASG), a modalidade da licitação, o número e o ano da licitação, a descrição do objeto e a descrição dos itens (materiais e serviços) da licitação, conforme ilustrado na figura 3.

A seção Serviços de Tecnologia da Informação e Comunicação do CATSER é composta por 7 divisões, 27 grupos, 43 classes e 202 categorias de serviços de TI (BRASIL, 2020). Ressalta-se que os serviços de TI não são divididos em subclasses, portanto esses serviços estão vinculados às classes diretamente, conforme ilustrado na figura 4.

Figura 3 | Exemplo de metadados associados à licitação

➔ DOWNLOAD DE EDITAIS

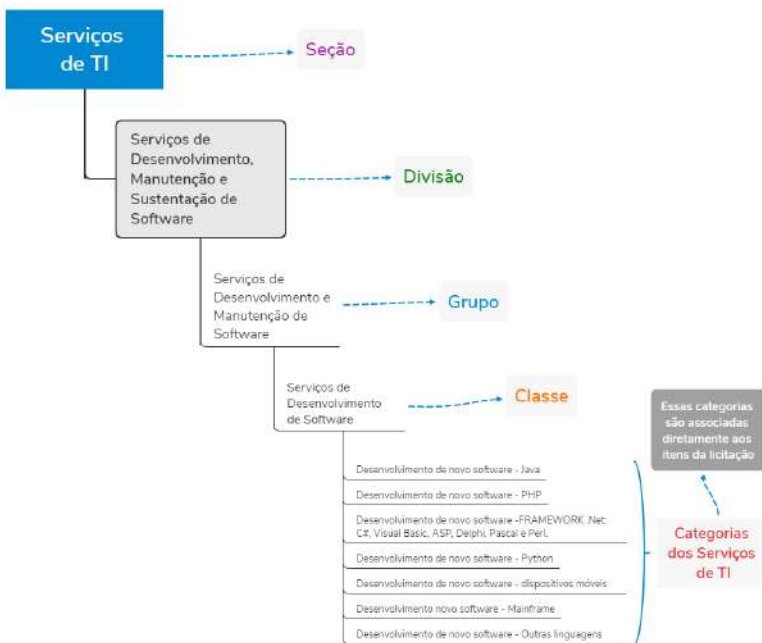
- MINISTÉRIO PÚBLICO DA UNIÃO
Ministério Público do Trabalho
PROCURADORIA GERAL DO TRABALHO
Código da UASG: 200200

Descrição da licitação

• **Pregão Eletrônico Nº 12/2021**
Objeto: Objeto: Pregão Eletrônico - Seleção de empresa especializada no fornecimento de subscrições Microsoft no regime Enterprise Agreement Subscription - EAS, buscando a renovação e expansão do licenciamento atualmente utilizado pelo MPT, de acordo com as especificações e condições descritas no edital e seus anexos.
Edital a partir de: 14/05/2021 das 08:00 às 11:59 Hs e das 12:00 às 17:59 Hs
Endereço: Saun Qd 5 Lote c Torre a - Asa Norte - BRASÍLIA (DF)
Telefone: (0xx61) 33148518
Fax: (0xx61)
Entrega da Proposta: 14/05/2021 às 08:00Hs

- **Itens de Serviços** **Categoria**
 - 1 - Software como Serviço - Saas**
 Office 365 E1_0365E1 ShrdSvr ALNG SubsVL MVL PerUsr. SKU T6A-00024. **Descrição do item da licitação**
 Tratamento Diferenciado: -
 Aplicabilidade Decreto 7174: -
 Aplicabilidade Margem de Preferência: Não
 Quantidade: 700
 Unidade de fornecimento: Unidade
 - 2 - Software como Serviço - Saas** **Categoria**
 Enterprise Mobility + Security E3 - Azure AD P1, Intune e Azure Information Protection, EntMobandSecE3Full ShrdSvr ALNG SubsVL MVL PerUsr. SKU AAA-10732. **Descrição do item da licitação**
 Tratamento Diferenciado: -
 Aplicabilidade Decreto 7174: -
 Aplicabilidade Margem de Preferência: Não
 Quantidade: 700
 Unidade de fornecimento: Unidade

Figura 4 | Exemplo da hierarquia das categorias dos serviços de TI



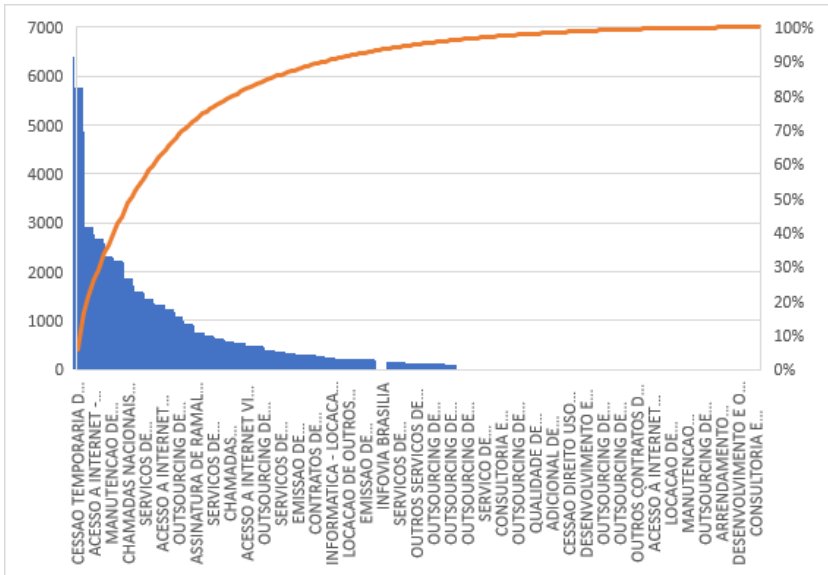
Em consulta realizada no DW-SIASG, extraíram-se 110.889 descrições de itens de licitação que foram associados a uma das 202 categorias do catálogo de serviços de TI entre 2017 e 2021.

As cinco categorias de serviços de TI com a maior quantidade de itens estão representadas no quadro 1. Verificou-se que 20% das categorias (41 de 202) representaram cerca de 77% dos itens (84963 de 110889), portanto os dados se aproximam do estabelecido pelo Princípio de Pareto, conforme figura 5. Trabalhos futuros poderão avaliar a possibilidade de eliminar as categorias menos utilizadas.

Quadro 1 | Categorias de serviços de TI com maior quantidade de itens

Categorias	Quant.	Perc.
CESSAO TEMPORARIA DE DIREITOS SOBRE PROGRAMAS DE COMPUTADOR LOCAÇÃO DE SOFTWARE	6400	5,77%
SERVICOS DE MANUTENCAO DE IMPRESSORAS, COPIADORAS, PLOTTER E SCANNERS	6308	5,69%
TELEFONIA FIXA COMUTADA / CONVENCIONAL	5771	5,20%
ACESSO A INTERNET VIA CABO	4871	4,39%
SERVICO DE LINK VIA CABO	3163	2,85%
Total	26513	23,90%

Figura 5 | Gráfico de Pareto com as categorias em ordem decrescente em frequência



PRÉ-PROCESSAMENTO DOS DADOS

Os itens de licitação dos últimos cinco anos (2017 a 2021) e categorizados manualmente pelos pregoeiros nas Categorias de Serviços de TI foram extraídos do banco de dados SIASG-DW, resultando em um total de 110.889 registros. A descrição do item de licitação, o código e o nome da categoria foram extraídos para um banco de dados no formato SQLite.

O banco de dados SQLite foi importado em script Python. Os textos da descrição de cada um dos 110.899 itens foram processados para retirar acentos, pontuação, caracteres especiais e *stopwords*. Adicionalmente, aplicou-se um algoritmo de *stemming* com o objetivo de padronizar e reduzir a quantidade de *features* do modelo de aprendizado supervisionado.

Cabe destacar que o pré-processamento não avaliou a utilização de algoritmos de correção ortográfica para melhorar o significado das descrições dos itens. Esse aprimoramento pode ser objeto de trabalhos futuros.

Adicionalmente, essa etapa não contemplou a verificação da exatidão da classificação manual dos itens de licitação pelos pregoeiros. Portanto, a quantidade de erros da classificação manual da base de dados é desconhecida e pode afetar o aprendizado supervisionado. Essa verificação do desempenho humano na tarefa de classificação da base de dados poderá ser objeto de trabalhos futuros, com o objetivo de melhorar o modelo. Os quadros 2 e 3 apresentam exemplos de itens de licitações associados incorretamente com categorias de TI.

Quadros 2 | Exemplos de itens de licitações que não se referem a serviços de TI associados a categorias de TI.

ID	Descrição da Categoria	Descrição do Item de Licitação
27510	CONTRATOS DE PRESTACAO DE SERVICOS DE ASSISTENCIA TECNICA E CIENTIFICA	CONTRATAÇÃO DE EMPRESA ESPECIALIZADA NA PRESTAÇÃO DE SERVIÇOS DE ASSISTÊNCIA TÉCNICA [...] PARA FOCOS CIRÚRGICOS DE TETO, INSTALADOS NAS UNIDADES DA REDE HOSPITALAR DA SECRETARIA [...].
27049	SERVICOS DE HOSPEDAGEM DE APLICATIVOS E PROGRAMAS	APARTAMENTO TRIPLO COM SUÍTE COM DIMENSÃO MÍNIMA DE 16,00 M2, CONTENDO NO MÍNIMO: 3 CAMAS [...].
26980	CENTRAL DE SERVICOS DE TIC	MANUTENÇÃO PREVENTIVA E CORRETIVA, POR MEIO DE REDE CREDENCIADA, PARA ATENDER AOS VEÍCULOS E ÀS MÁQUINAS DA FROTA DO MUNICÍPIO DE [...]. INCLUINDO: O FORNECIMENTO DE PNEUS [...].

Quadro 3 | Exemplos de itens de licitações que se referem a serviços de TI associados a categorias de TI erradas.

ID	Descrição da Categoria	Descrição do Item de Licitação
24333	SERVICO DE LICENCA PELO USO DE SOFTWARE	CERTIFICADO DIGITAL PARA PESSOA FÍSICA TIPO A3 COM VALIDADE DE TRÊS ANOS SEM FORNECIMENTO DE TOKEN.
25860	DESENVOLVIMENTO E OU EVOLUCAO DE SOFTWARE - PHP	SERVIÇO E NUVEM (INFRAESTRUTURA) AZURE PREPAYMENT
26018	DOCUMENTACAO DE SOFTWARE	FORNECIMENTO DE LICENÇAS DE USO DO SOFTWARE MICROSOFT SQL SERVER , NA MODALIDADE DE LICENCIAMENTO PERPÉTUO, DESTINADAS [...].
26980	CENTRAL DE SERVICOS DE TIC	CONTRATAÇÃO DE SERVIÇOS DE SUPORTE TÉCNICO NA PLATAFORMA ORACLE PRIMAVERA P6 [...].

MODELAGEM

A partir de novos itens de licitação, o modelo deve apontar uma ou mais categorias prováveis. Portanto, essa etapa consistiu na seleção e na análise dos algoritmos de aprendizagem de máquina supervisionado multiclasse. Neste estudo, experimentaram-se os algoritmos LinearSVC, Multinomial Naive Bayes, Logistic Regression, MLP e SGD.

Considerando que esses algoritmos esperam *features* (recursos/dimensões) numéricas, existe a necessidade de empregar algoritmos de vetorização textual. Cada documento foi representado por um vetor binário do tamanho do vocabulário de todo conjunto de documentos. O algoritmo de vetorização textual aplicado neste estudo foi o TF-IDF, que visa calcular a medida de concentração de determinada palavra em relativamente poucos documentos.

O vocabulário do conjunto de documentos resultou em 12.640 palavras únicas (unigramas). Com o objetivo de melhorar o desempenho, além de considerar unigramas, o modelo foi parametrizado para considerar também

os bigramas (combinação de duas palavras) no vocabulário do modelo. Assim, o vocabulário resultou em 55.395 *tokens* (unigramas e bigramas).

Inicialmente, o modelo deveria considerar 110.889 itens de licitação distribuídos em 202 categorias de serviços de TI. Entretanto, algumas categorias possuíam poucos itens de licitação, sendo a menor delas com apenas 3 registros. Assim, optou-se por considerar apenas categorias com no mínimo 50 registros, reduzindo o quantitativo para 109.793 itens de licitação distribuídos em 153 categorias.

Considerando o quantitativo de itens de licitação e o vocabulário, a matriz TF-IDF apresentou o tamanho de 109.889×55.395 . Cada elemento da matriz ocupava 8 bytes. Portanto, a matriz completa ocupou cerca de 48,6 gigabytes em memória.

O modelo foi construído a partir do treinamento de 80% dos itens de licitação escolhidos aleatoriamente. Os 20% restantes do conjunto de dados foram reservados para testar a efetividade do algoritmo.

Entre as métricas utilizadas para a avaliação de modelos de aprendizagem supervisionada, destaca-se a acurácia pela simplicidade de cálculo e interpretação. A fórmula acurácia consiste na quantidade de registros corretamente previstos dividido pela quantidade total de registros.

Os algoritmos de aprendizagem supervisionada calculam a probabilidade de todas as categorias para cada nova entrada de dados (descrições de itens de licitação neste estudo) pertencer a determinada categoria. Assim, o algoritmo pode apresentar como resultado as K categorias de maior probabilidade, sendo K no mínimo 1 e no máximo a quantidade de categorias do modelo.

Ao aumentar o conjunto de categorias previstas que será apresentado ao usuário, o modelo aumenta a acurácia. Por exemplo, a probabilidade de um modelo acertar a categoria correta ao apresentar uma lista de dez categorias é maior do que a probabilidade de apresentar apenas uma categoria. No limite, quando o modelo apresenta o conjunto de todas as categorias para o usuário, a probabilidade de acerto é 100%.

Para exemplificar, o quadro 4 apresenta as cinco categorias mais prováveis, calculada pelo modelo, para a descrição do item de licitação: “CRÉDITOS UNIVERSAIS (UNIVERSAL CREDITS) NA QUANTIDADE SUFICIENTE PARA HABILITAR 60 (SESSENTA) OCPUS [...]”.

Quadro 4 | Exemplo da lista com as cinco categorias com maior probabilidade de serem a categoria correta de determinado item de licitação.

Categoria	Probab.
INFRAESTRUTURA COMO SERVICO - IAAS	0,025682
PLATAFORMA COMO SERVICO - PAAS	0,009313
OUTROS SERVICOS DE GERENCIAMENTO DE INFRAESTRUTURA DE TECNOLOGIA DA INFORMACAO E COMUNICACAO (TIC)	0,009239
SERVICOS DE CONSULTORIA EM TECNOLOGIA DA INFORMACAO E COMUNICACAO (TIC)	0,008426
MANUTENCAO EVOLUTIVA DE SOFTWARE (ACRESCIMO DE NOVAS FUNCIONALIDADES) - OUTRAS LINGUAGENS	0,008388

Cabe destacar que a apresentação de um pequeno conjunto de categorias (ex.: 10 categorias) para o usuário escolher, ainda assim, atenderia o objetivo do trabalho no sentido em que diminuiria a complexidade da tarefa do usuário em escolher milhares de categorias (atualmente existem 141.626 categorias de materiais e 2.683 categorias de serviços).

Assim, esse modelo considerou a métrica “Acurácia Top K”. Caso a categoria prevista esteja entre as K categorias mais prováveis, essa métrica considera que o modelo previu a categoria corretamente. No exemplo do quadro 4, a métrica “Acurácia Top 5” consideraria que o modelo previu corretamente a categoria se ela estivesse presente nessa lista, em qualquer posição.

EXPERIMENTOS E RESULTADOS

O modelo multiclasse proposto contém 153 categorias. Foram utilizados 109.793 registros (itens de licitação) para treinar (80%) e testar (20%) o modelo.

O vocabulário do modelo apresentou 55.395 *tokens*. Conforme exposto, a matriz TF-IDF ocupou 48,6 gigabytes de memória.

Verifica-se que há razoável requisito de memória apenas para o treinamento do modelo para prever as categorias dos serviços de TI. Em um exercício de extrapolação, verifica-se a existência de 1.740.167 descrições de itens de licitação para todos os tipos de serviço (incluindo os serviços não relacionados com TI) e 4.960.953 descrições de itens de licitação para todos os tipos de materiais, licitados entre 2017 e 2021. Considerando que o vocabulário seja pelo menos 55.395 *tokens*, ou seja, o mesmo do modelo atual, ter-se-ia um requisito de memória de cerca de 771 gigabytes e 2.198 gigabytes para as matrizes com a descrição de itens de licitação associados ao CATSER e ao CATMAT, respectivamente.

Para este trabalho, considerou-se como aceitável recomendar entre uma e dez categorias para o usuário após a inserção de nova descrição de item de licitação. Cabe destacar que, mesmo apresentando uma lista de dez categorias ao usuário, reduz-se a complexidade da tarefa e diminui a possibilidade de erros de classificação, pois a escolha do usuário será restrita em um conjunto pequeno frente às 153 categorias neste estudo.

Considerando que o modelo pode apresentar uma lista com até dez categorias, o desempenho do modelo foi medido pela métrica Acurácia “Top K”. Utilizou-se o método *top_k_accuracy_score* da biblioteca *sklearn* que calcula a acurácia considerando o número de vezes que a categoria correta está entre as *K* categorias previstas com maior probabilidade (SKLEARN.METRICS.TOP_K_ACCURACY_SCORE, 2022).

O quadro 5 apresenta o resultado da métrica de acurácia “Top K” em algoritmos de aprendizagem supervisionada com diferentes tamanhos de lista a serem apresentados para o usuário. Por exemplo, a métrica acurácia Top 1 verifica se a única categoria apresentada ao usuário é igual a categoria rotulada. Já a métrica acurácia Top 5 considera que o algoritmo previu corretamente se a categoria estiver entre as cinco categorias apresentadas ao usuário.

Quadro 5 | Acurácia Top K dos algoritmos de classificação testados.

Algoritmo	Acurácia			
	Top 1	Top 3	Top 5	Top 10
LinearSVC (SVM)	74,7%	86,6%	89,8%	92,9%
MultinomialNB	55,0%	75,3%	81,6%	87,6%
LogisticRegression	69,7%	84,2%	89,1%	93,7%
MLP	74,0%	86,4%	90,2%	94,0%
SGD	73,0%	86,5%	90,4%	94,5%

Verifica-se que o algoritmo LinearSVC (*Scalable Linear Support Vector Machine*) apresentou o melhor desempenho na acurácia Top 1 e Top 3. Esse algoritmo é similar ao algoritmo SVC (*Support Vector Machine*), mas possui maior escalabilidade com amostras grandes (SKLEARN.SVM.LINEARSVC, 2022). O melhor desempenho na acurácia Top 5 e na Top 10 foi manifestado pelo algoritmo SGD (*Stochastic Gradient Descent*).

Embora este estudo não tenha analisado a taxa de acurácia da categorização pelo usuário (desempenho humano), as altas taxas de acurácia (entre 74,7% e 94,5%, ao recomendar entre uma e dez categorias) indicam a viabilidade do modelo. Conseqüentemente, sugere-se a ampliação deste estudo para desenvolver um sistema de recomendação, incluindo todas as categorias do catálogo CATMAT/CATSER, integrado ao sistema Compras.gov.br.

CONCLUSÃO

O estudo apresentado envolveu a utilização de mineração de textos e aprendizagem de máquina supervisionado para desenvolver um sistema de recomendação das categorias do CATMAT/CATSER, a partir da descrição de um item de compra. A prova de conceito elaborada, no escopo de Serviços de TI, demonstrou a viabilidade para ser implementada no sistema

Compras.gov.br, evidenciando que, ao apresentar dez categorias, o algoritmo apresenta a categoria correta em 94,5% das previsões.

O estudo demonstrou o contexto do negócio dos catálogos CATMAT/CATSER dentro do sistema Compras.gov.br e sua importância para a administração pública: auxiliar na identificação de preços praticados para balizar o valor estimado da licitação. O levantamento de dados indicou que a quantidade de categorias dos catálogos CATMAT e CATSER é extenso, contendo 141.626 e 2.683 categorias respectivamente.

A etapa de pré-processamento foi executada com o objetivo de preparar os textos para serem utilizados pelos algoritmos de aprendizado de máquina. A modelagem representou os textos em uma forma em que os algoritmos de classificação fossem capazes de interpretar mediante o algoritmo de vetorização textual TF-IDF. A etapa de experimentos e resultados apresentou o resultado das principais métricas dos algoritmos de classificação: LinearSVC (SVM), MultinomialNB, LogisticRegression, MLP e SGD.

A acurácia Top 1 do algoritmo LinearSVC (SVM) apresentou o melhor desempenho de classificação: 74,7%. Esse resultado indica a taxa de acerto, considerando que o algoritmo apresenta apenas uma categoria prevista. Já a acurácia Top 10 do algoritmo SGD demonstrou o melhor resultado. Assim, ao apresentar dez categorias para o usuário, a taxa de acerto desse algoritmo resulta em 94,5% das categorias previstas.

Considerando os resultados expostos, a implantação de um sistema de recomendações das categorias para o CATMAT/CATSER é apropriada. Esse sistema pode reduzir as dezenas de milhares de opções do catálogo para uma dezena de opções de categorias para o usuário.

Destaca-se que, com a implementação desse modelo no sistema Compras.gov.br, a economia de recursos públicos ocorrerá em duas perspectivas. A primeira é a diminuição da carga de trabalho dos usuários do sistema, devido a maior celeridade na escolha das categorias dos itens de licitação. A segunda é a redução de sobrepreços em licitações devido às estimativas de preços aceitáveis com maior exatidão. Com a redução dos erros humanos na

classificação de itens de licitações, haverá maior facilidade em encontrar os reais preços praticados no contexto da Administração Pública.

REFERÊNCIAS

BRASIL. Secretaria de Gestão do Ministério da Economia. **Instrução Normativa SEGES/ME n. 65/2021**. Disponível em: <<https://www.gov.br/compras/pt-br/aceso-a-informacao/legislacao/instrucoes-normativas/instrucao-normativa-seges-me-no-65-de-7-de-julho-de-2021>>. Acesso em: 30 set. 2022.

_____. Secretaria de Gestão Pública do Ministério da Economia. **Planilha vigente do Catálogo de Materiais (CATMAT) e Catálogo de Serviços (CATSER)**. Disponível em: <<https://www.gov.br/compras/pt-br/aceso-a-informacao/consulta-detalhada/planilha-catmat-catser>>. Acesso em: 30 set. 2022.

_____. Secretaria de Gestão Pública do Ministério da Economia. **CATSER - Catálogo de Serviços e CATMAT - Catálogo de Materiais**. 2020. Disponível em: <<https://www.gov.br/compras/pt-br/aceso-a-informacao/manuais/manual-siasg/manual-siasg-catalogo/manual-catmat-e-catser-2020.pdf>>. Acesso em: 30 set. 2022.

_____. Secretaria de Gestão Pública do Ministério da Economia. **Conheça o Compras.gov.br**. Disponível em: <<https://www.gov.br/compras/pt-br/sistemas/conheca-o-compras>>. Acesso em: 30 set. de 2022.

FERNANDES, C. C. C. **Compras públicas: da reforma gerencial à inovação** impulsionada pela tecnologia. Brasília: [s.n.].

MAHESH, B. **Machine learning algorithms-a review**. International Journal of Science and Research (IJSR). [Internet], v. 9, p. 381–386, 2020.

NASTESKI, V. **An overview of the supervised machine learning methods**. Horizons. b, v. 4, p. 51–62, 2017.

RIBEIRO, C. G.; INÁCIO JÚNIOR, E. **O Mercado de compras governamentais brasileiro (2006-2017): mensuração e análise.** Brasília, 2019.

SEN, P. C.; HAJRA, M.; GHOSH, M. Supervised classification algorithms in machine learning: A survey and review. In: **Emerging technology in modelling and graphics.** [s.l.] Springer, 2020.

SKLEARN.METRICS.TOP_K_ACCURACY_SCORE: scikit-learn 1.1.2 documentation. 2022. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.top_k_accuracy_score.html>. Acesso em: 30 set. 2022.

SKLEARN.SVM.LINEARSVC: scikit-learn 1.1.2 documentation. 2022. Disponível em: <<https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html>>. Acesso em: 30 set. 2022.

SOYLU, A. *et al.* **Data Quality Barriers for Transparency in Public Procurement.** Information, v. 13, n. 2, p. 99, 2022.



TEMA: BANCOS PÚBLICOS

FERRAMENTA: ESTATÍSTICA
INFERENCIAL, APRENDIZADO DE
MÁQUINA SUPERVISIONADO

AVALIAÇÃO DA ATUAÇÃO DOS BANCOS PÚBLICOS COMO AGENTES DE POLÍTICA PÚBLICA NA COBERTURA DE FALHAS DE MERCADO¹

MÔNICA RONDINA⁽¹⁰⁾

INTRODUÇÃO

As instituições financeiras atuam no mercado intermediando o uso de moeda. Captam depósitos, poupança e aplicações de seus clientes e emprestam-nos a outros clientes que necessitam de recursos, financiando, assim, suas atividades produtivas. Ao concederem empréstimos e financiamentos, tendem a buscar a melhor equação, alocando seus recursos onde houver maior rentabilidade e menor risco (QUAGLIIO, 2014). Com isso, formam-se lacunas de mercado, uma vez que há projetos que demandam financiamento, mas não interessam às instituições privadas: projetos em regiões pobres

1 Este artigo resulta de trabalho de conclusão orientado pelo Professor Francisco Eduardo de Holanda Bessa.

ou distantes, clientes de baixa renda, micro e pequenas empresas, projetos novos e sem garantias, projetos com risco elevado, projetos que demandam longo prazo de maturação ou grande volume de investimentos (LOBATO, 2020; THORTON, 2011 *apud* SOIHET; CABRAL, 2016).

O presente trabalho buscou identificar se os bancos públicos brasileiros evidenciarão maior disposição, enquanto instrumentos de execução de política de crédito, para oferecer financiamento a setores da economia para os quais os bancos privados teriam menor propensão para emprestar.

Os bancos públicos brasileiros atuam no cumprimento de políticas públicas, das seguintes formas: 1) oferta de crédito a setores que não são de interesse dos bancos privados 2) estímulo ao desenvolvimento regional; 3) atuação anticíclica, expandindo a liquidez em momentos de crise; e 4) promoção da inclusão bancária (ARAUJO, 2011).

O mercado de crédito possui lacunas em segmentos que, embora tenham importância social, podem não trazer o retorno esperado pelas instituições financeiras, em particular exigem-se financiamento de longo prazo (STIGLITZ, 1994, *apud* HORN; FEIL, 2019). Segundo Gonçalves (2009, *apud* RAMOS; MARTHA JUNIOR 2010), o Estado atua corrigindo ou eliminando imperfeições que afetam o mercado.

O tema reveste-se de especial importância, dado que a ausência ou a ineficiência de uma política de crédito teria grande impacto negativo às populações menos favorecidas. A confirmação do papel dos bancos públicos pode orientar melhor atuação do governo no exercício de seu papel de “dono” na governança dessas instituições financeiras federais. Acrescenta-se que os resultados podem contribuir para reafirmar ou para repensar o papel dos bancos públicos enquanto agentes da política creditícia governamental.

O trabalho, ainda, buscou avaliar a oferta de crédito a setores específicos da economia, que justificassem a atuação governamental por meio de políticas públicas. Foram avaliados os financiamentos concedidos a atividades socialmente relevantes, tradicionalmente atendidas pela política creditícia no Brasil: os setores rural e agroindustrial, de infraestrutura e habitacional.

Políticas de apoio à atividade rural buscam evitar que condições adversas, como risco climático, baixa elasticidade-preço e elasticidade-renda afetem significativamente a rentabilidade do produtor rural e a oferta de produtos agrícolas (RAMOS; MARTHA JUNIOR, 2010). Araújo (2011) acrescenta que a política de crédito rural teria também cumprido, em anos passados, a função de compensação ao setor agropecuário por outras políticas de intervenção nos mercados agrícolas. Leite e Junior (2014) creditam a expansão da atividade agrícola no país à política de financiamento rural e agroindustrial, com crescente aporte de recursos públicos. Para os autores, a contribuição do setor público mostra-se relevante, ainda que o financiamento de grandes propriedades ocorra concomitantemente a partir de fontes governamentais e privadas.

Monteiro e Veras (2017) relacionam políticas habitacionais — incluindo o crédito à habitação — à redução de desigualdades sociais e à promoção do desenvolvimento humano. Cunha (2020) pontua a dificuldade em combater o déficit habitacional brasileiro por meio de uma política pública, uma vez que tal déficit está fortemente associado à elevada concentração de renda. Ferreira *et al.* (2019) destacou limitações do programa *Minha Casa Minha Vida* na redução do déficit habitacional, ainda que tal programa tenha contribuído para evitar o seu agravamento.

No que tange ao financiamento à infraestrutura, a função social de uma política creditícia ao segmento é menos óbvia. Para Turolla *et al.* (2013), é preciso avaliar a relação custo e benefício na assunção de grandes riscos fiscais decorrentes do financiamento público ao setor. Esses autores destacam ser necessário investigar se o financiamento público induziria maior participação do capital privado. Campos (2021) menciona um modelo de financiamento híbrido para o setor, com a participação de agentes financeiros públicos em conjunto com o setor privado.

DESENVOLVIMENTO

O sistema financeiro brasileiro encontra-se em um estágio inicial de maturidade, existindo modalidades de crédito e de serviços financeiros que o setor privado não tem interesse em oferecer em determinadas regiões ou segmentos econômicos. Nesse sentido, os bancos públicos cumprem a função de suprir falhas de mercado, além de serem instrumentos da política econômica governamental, como reguladores da concorrência ou atuando para minimizar os efeitos de ciclos econômicos (LOBATO, 2020).

Quando há incerteza, como nos períodos de crise, as instituições financeiras tendem a priorizar a sua própria liquidez, reduzindo financiamentos em que são percebidos maiores riscos ou menores rentabilidades (QUAGLIO, 2014; CROCCO; NOGUEIRA; ANDRADE; DINIZ, 2011).

Governos atuam no sistema financeiro por meio de bancos de desenvolvimento. Com o auxílio de bancos públicos, podem implementar uma política creditícia que atenda às falhas de mercado, uma vez que o setor privado não teria disposição de atender de forma adequada aos projetos de risco elevado como aqueles situados em regiões pobres ou distantes, pequenos, sem garantias, novos, com novas tecnologias, que exijam investimentos elevados ou de longo prazo (TORTON, 2011 apud SOIHET; CABRAL, 2016).

Henriques (2020) alerta para a necessidade de que os prazos de financiamento sejam conciliados com os horizontes dos projetos, sendo necessário haver mecanismos capazes de mobilizar fundos com prazos compatíveis. Em países capitalistas, a principal solução para viabilizar aportes de recursos a projetos é provida pelo mercado de capitais, mas nem todas as economias têm a maturidade para viabilizar a intermediação de recursos poupados por essa via, em favor dos projetos que demandam os aportes. Nesses casos, a solução de financiamento a projetos de baixa atratividade depende da atuação de bancos públicos. Além disso, há países que contam com esses bancos como instrumento para aplicação de diretrizes de desenvolvimento.

Mettenheim (2015) sugere que bancos públicos teriam uma vantagem competitiva sobre bancos privados, ao terem acesso a capital barato e menores pressões por lucros. Os governos, como investidores, teriam maior grau de aversão ao risco e horizontes mais longos que os investidores privados, o que também conferiria vantagem aos bancos públicos.

Apesar do importante papel que os bancos públicos têm desempenhado na concessão de crédito a setores específicos, ainda não são capazes de suprir plenamente a demanda, havendo a necessidade de atuação compartilhada com instituições privadas, sobretudo com outros agentes financeiros de longo prazo. No Brasil, há uma tendência de crescimento do crédito em proporção ao produto interno bruto (PIB), com significativa participação de bancos públicos, muito embora os estoques de crédito dos bancos privados venham crescendo a taxas mais elevadas (ARAÚJO; CINTRA, 2011).

Por fim, importa destacar que a aplicação de crédito no Brasil não necessariamente tem sido efetiva na mitigação de diferenças econômicas regionais. A região Sudeste concentra a maior parte das operações de crédito. Apesar do crescimento da oferta de crédito nas regiões Norte e Nordeste, a sua participação ainda é baixa (QUAGLIO, 2014). O sistema financeiro brasileiro é formado majoritariamente por instituições de abrangência nacional. Não obstante, os bancos adotam estratégias diferenciadas no território, contribuindo para a ampliação das disparidades regionais (CROCCO; NOGUEIRA; ANDRADE; DINIZ, 2011).

Além disso, uma avaliação realizada no âmbito do Conselho de Monitoramento e Avaliação de Políticas Públicas da operacionalização dos Fundos Constitucionais de Financiamento do Norte (FNO), Nordeste (FNE) e Centro-Oeste (FCO) observou baixo grau de correlação entre o volume de financiamentos dos fundos e os indicadores de desenvolvimento econômico das regiões beneficiadas (CMAP, 2021).

BANCOS PÚBLICOS FEDERAIS BRASILEIROS

Cavalcante (2002, *apud* SILVA *et al.*, 2016) define o sistema financeiro como “um conjunto de instituições e instrumentos financeiros que possibilita a transferência de recursos entre ofertantes e tomadores de crédito”. A Lei n. 4.595/64 define as instituições financeiras públicas como órgãos auxiliares da execução da política de crédito do governo federal e atribui ao Conselho Monetário Nacional (CMN) a função de orientar a aplicação das instituições públicas e privadas de forma a propiciar o desenvolvimento econômico.

No Brasil há três grandes bancos públicos federais, que tradicionalmente atuam como agentes de políticas de crédito em segmentos específicos, além de dois bancos voltados ao desenvolvimento regional. O Banco Nacional do Desenvolvimento (BNDES) é o principal instrumento de execução da política de financiamento ao investimento do Governo Federal (Lei n. 4.595/64). Suas principais fontes de recursos são o Fundo de Amparo ao Trabalhador (FAT) e o Tesouro Nacional, mas também opera outros fundos governamentais, como o Fundo de Investimento do Fundo de Garantia por Tempo de Serviço (FI-FGTS), além de dispor de recursos de outras fontes (BNDES, 2021).

O Banco do Brasil (BB) executa políticas públicas com recursos governamentais provenientes de fundos de financiamento e com recursos próprios. Os principais programas sob sua gestão destinam-se ao estímulo da agricultura familiar, à concessão de crédito a produtores rurais, ao financiamento à exportação de bens e serviços e à concessão de microcrédito. O banco é também responsável pela aplicação de recursos do FCO (Fundo Constitucional de Financiamento do Centro-Oeste), destinados ao crédito rural e aos microempreendedores individuais e à micro e pequenas empresas (BANCO DO BRASIL, 2021).

A Caixa Econômica Federal (Caixa) opera como instituição financeira e como mandatária da União em diversos produtos e programas governamentais voltados à promoção de políticas públicas. No financiamento à habitação, a Caixa aplica recursos do Sistema Brasileiro de Poupança e Empréstimo (SBPE), FGTS e Programas Minha Casa Minha Vida (PMCMV), além do Programa Casa

Verde Amarela (PCVA). Também opera linhas de crédito para o Setor Público, por meio dos projetos de saneamento, geração de energia, iluminação pública, pavimentação, construção de escolas, creches e hospitais. Além disso, atua como agente operador do Governo Federal para políticas públicas de desenvolvimento social e urbano do Brasil, a partir dos recursos provenientes do Orçamento Geral da União (OGU), além de ser agente operador e de pagamento de diversos benefícios sociais (CAIXA, 2021).

O Banco da Amazônia (Basa) é um agente do Governo Federal na implementação de políticas públicas voltadas ao desenvolvimento econômico e social da Amazônia. Promove o fomento do desenvolvimento da região com a utilização de recursos do Fundo Constitucional de Financiamento do Norte (FNO) e outras fontes. Seus principais produtos de crédito são voltados ao microcrédito, ao financiamento da agricultura familiar, à infraestrutura, à micro e pequenas empresas e ao agronegócio (BASA, 2022).

O Banco do Nordeste (BNB) foi criado para atuar no chamado Polígono das Secas, porção do território brasileiro atingida periodicamente por estiagens prolongadas. Ele administra o Fundo Constitucional de Financiamento do Nordeste (FNE), além de operar o Fundo de Desenvolvimento do Nordeste (FDNE) e a carteira de títulos do Fundo de Investimentos do Nordeste (FINOR). É também agente de crédito do PRONAF (BNB, 2021).

BANCOS ESTADUAIS E AGÊNCIAS DE FOMENTO

A partir dos anos 1960, estimuladas pela criação da Finame, muitas unidades federativas constituíram suas próprias instituições públicas de fomento. Todavia, nos anos 1990, a maior parte delas foi extinta ou privatizada, enquanto outras deram origem a agências de fomento, criadas por meio da Medida Provisória n. 1514/96. São instituições financeiras não bancárias, sob controle de unidade da federação, importantes para a operacionalização de políticas públicas de âmbito local, embora sejam pequenas em relação ao conjunto do sistema financeiro, mesmo nas economias estaduais onde atuam (HORN; FEIL, 2019).

ARCABOUÇO PRUDENCIAL

Na década de 1970, para minimização de riscos relacionados aos impactos causados por crises em bancos, foi criado o *Comitê de Supervisão Bancária de Basileia*, originalmente composto por 12 países, e em 1988 veio a ser firmado o acordo de Basileia, que estabeleceu a necessidade de um aporte de capital mínimo aos ativos das instituições financeiras, ponderados pelo risco (RÊGO, 1995).

No Brasil, as normas do acordo de Basileia foram incorporadas com a edição, pelo Bacen, da Resolução n. 2.099/1994, que estabeleceu padrões de capital e patrimônio líquido ajustado, compatível com o grau de risco da estrutura de seus ativos, passivos e contas de compensação, a serem observados pelas instituições financeiras, podendo ser considerados como tal depósitos feitos em conta vinculada específica custodiada no Bacen (BCB, 1994).

Em 2004, foi publicado pelo Comitê de Basileia o documento denominado *Novo Acordo de Capital da Basileia*, também denominado Basileia II. Em âmbito nacional, os dispositivos de Basileia II foram regulamentados pelo CMN e publicados por meio da Resolução BCB n. 3.490/2007. Foi exigido das instituições financeiras um patrimônio de referência composto por diversas parcelas de risco pelos quais os ativos deveriam ser ponderados (BCB, 2007).

Com o advento da crise financeira de 2008 e diante da internacionalização bancária e crescente integração dos mercados financeiros, seguida de instabilidade dos sistemas monetários nacionais, o modelo de regulação financeira internacional precisou ser revisto, de modo a inibir práticas arriscadas adotadas pelos bancos. O cálculo do risco de mercado havia sido deixado à discricionariedade dos próprios bancos, o que impôs aos reguladores a necessidade de passar a aprovar e validar essas apurações. Não foram incomuns situações em que os reguladores foram “intimidados” pelas instituições com uma avalanche de dados, modelos e simulações. Além disso, a percepção de maiores riscos durante um período recessivo levava à reclassificação das carteiras, elevando o volume de capital mínimo requerido e inevitavelmente à redução do volume de crédito concedido, causando uma atuação pró-cíclica.

Assim, foi processado um conjunto de ajustes denominado Basileia III, que teve por objetivo mobilizar a criação de um colchão de conservação de capital e um colchão destinado a suportar efeitos de ciclos econômicos. Também avançou ao aumentar a importância dos processos de supervisão das instituições e transparência e ao introduzir medidas para assegurar um grau satisfatório de liquidez das instituições, no curto e no longo prazo (LEITE; REIS, 2013). No Brasil, a adoção de Basileia III concretizou-se com a edição das Resoluções de n. 4.192 a 4.195 do CMN, todas publicadas em 1º de março de 2013, e complementado pelas Circulares de n. 3.634 a 3.648, do BCB, divulgadas em 4 de março de 2013.

A Resolução BCB n. 4195/2013 introduziu o conceito de conglomerado prudencial, abrangendo, além das instituições financeiras e demais instituições autorizadas a funcionar pelo BCB, administradoras de consórcio, sociedades de crédito, emissoras de cartões e seguradoras (BCB, 2013).

METODOLOGIA

Foi realizada uma pesquisa do tipo quantitativa, a partir de dados disponíveis em uma base divulgada pelo BCB, em [*IF.data \(bcb.gov.br\)*](http://IF.data(bcb.gov.br)), contendo informações sobre instituições por ele autorizadas a funcionar. Nessas análises, foram utilizados dados das carteiras ativas de crédito concedidas a pessoas físicas e jurídicas, bem como dados contábeis das instituições financeiras e dos conglomerados.

Com o auxílio da ferramenta Pandas, usando a linguagem de programação *Python*, no ambiente *Google Colab*, os dados foram consolidados em dois grandes *dataframes*, um deles com dados de conglomerados bancários e instituições financeiras independentes e o segundo com dados de conglomerados prudenciais. Foram então realizadas operações entre as variáveis disponíveis, sendo criadas, assim, novas variáveis. Essas operações foram realizadas, computacionalmente, para todos os registros dos *dataframes*.

A partir das informações contábeis fornecidas pelas instituições, foram calculados os indicadores: margem financeira bruta, margem financeira dos ativos, custo médio de captação, retorno médio das operações de crédito, *spread* médio, índice de eficiência, imobilização do patrimônio líquido e índice de perdas estimadas nas operações de crédito.

Para realização das análises, foram realizados, com o uso das ferramentas mencionadas, filtros e agrupamentos dos dados, conforme a necessidade. Foram expurgados registros nulos e, para algumas operações, registros com saldo zero, em especial quando a variável comporia o denominador de uma operação.

Para comparação entre variáveis no *dataframe* que continham dados das operações de crédito, foi utilizado o teste qui-quadrado, com o auxílio da ferramenta *scipy.stats*. Para realização desse teste, variáveis contínuas foram transformadas em variáveis categóricas, por meio do agrupamento dos valores por faixas. Do *dataframe* que continha dados contábeis dos conglomerados prudenciais, por sua vez, foi construído um modelo de regressão linear, com o uso das ferramentas *sklearn* e *statsmodels*. Nesse modelo, para identificação das variáveis que mais contribuem para explicar as diferenças entre instituições públicas e privadas, foi utilizada a ferramenta *Radom Forest Regressor*.

RESULTADOS E CONCLUSÃO

Os resultados indicaram que os bancos públicos comprometem uma maior proporção de suas carteiras com o financiamento à infraestrutura e à habitação em relação aos bancos privados. A carteira habitacional em relação à carteira total é maior nos bancos públicos, enquanto a proporção entre o saldo das carteiras de crédito rural em relação à carteira total é muito semelhante entre bancos públicos e privados. Uma possível explicação para esse resultado pode estar associada à efetividade do mecanismo de direcionamento obrigatório de crédito, que estaria de fato induzindo os bancos privados a se interessarem em atender o segmento rural.

As carteiras de crédito rural e habitacional das instituições bancárias apresentam crescimento constante ao longo dos anos, mais intensamente na carteira habitacional. Já o financiamento à infraestrutura vinha crescendo de forma significativa até 2015, a partir de quando passou a declinar, estabilizando-se a partir de 2020. Bancos públicos são responsáveis pelo maior volume de recursos investidos nesses setores e ainda são os principais indutores do seu crescimento.

Os dados indicam, ainda, que bancos públicos estariam mais dispostos a assumir riscos que bancos privados, sendo que as operações de crédito de maior risco são essencialmente atendidas por bancos públicos.

É esperado que o atendimento a clientes de maior risco possa levar a uma maior inadimplência. Assim, esse aspecto também foi objeto de avaliação. Os resultados confirmam que os bancos públicos assumem um maior risco, pois a inadimplência de suas carteiras é, de fato, maior.

Outra função social avaliada foi o atendimento à micro e pequenas empresas, das quais os bancos públicos são os maiores financiadores, embora sejam também os que mais fomentam grandes empresas, o que é um indicativo de uma maior capacidade de prover crédito de grande vulto.

Bancos públicos e privados vêm experimentando crescimento na quantidade de clientes ao longo dos anos. Nos bancos públicos, esse volume cresceu a taxas maiores, contribuindo para a inclusão bancária da população. Não obstante, o crescimento da quantidade de clientes não foi acompanhado de aumento na quantidade de operações de crédito.

Observa-se que o saldo médio das operações de crédito é menor nos bancos públicos, indicando que esses concedem créditos de valores menores, mais pulverizados, o que é também socialmente relevante. Além disso, bancos públicos operam financiamentos de mais longo prazo, especialmente, ao crédito rural e à infraestrutura.

Na vertente prudencial, observa-se que o Índice de Basileia dos bancos privados é sensivelmente superior ao dos bancos públicos. Os bancos públicos

detêm um patrimônio de referência muito superior ao dos bancos privados, o que era esperado em razão do seu porte e da origem do seu capital. Seus ativos ponderados pelo risco são também muito superiores, como consequência do atendimento à sua função pública.

Do exame das demonstrações contábeis dos bancos públicos federais, em particular das demonstrações da mutação do patrimônio líquido (DMPL), não foram observados aportes provenientes de recursos fiscais no período examinado (2013 a 2021). Os aumentos de capital observados no período decorreram, majoritariamente, da capitalização de reservas. Assim, observa-se que essas instituições estão sendo capazes de cumprir funções de políticas públicas e, ao mesmo tempo, de manter um crescimento sustentado a partir de seus próprios resultados.

Por fim, identificou-se que, de um conjunto de variáveis contábeis ou relativas ao porte das instituições, as que melhor explicam as diferenças entre instituições públicas e privadas são imobilização do patrimônio líquido, tipo de consolidação (se conglomerado com ou sem banco comercial, cooperativa de crédito singular, central de cooperativas, banco de desenvolvimento ou instituições não bancárias ou de pagamento), número de agências e passivo financeiro.

REFERÊNCIAS

ARAUJO, V. L. D; CINTRA, M. A. M. **O papel dos bancos públicos federais na economia brasileira**. 2011. Texto para discussão n. 1604, Instituto de Pesquisa Econômica Aplicada (Ipea).

ARAUJO, P.F.C. **Política de crédito rural**: reflexões sobre a experiência brasileira. DF: CEPAL. Escritório no Brasil/IPEA, 2011. (Textos para Discussão CEPAL-IPEA, 37).

BARROS, L.A.B; SILVA, C.K.S; OLIVEIRA, R.F.O. **Presença Estatal no Mercado de Crédito**: o papel dos bancos públicos e do crédito direcionado na crise de

2008. Brasília, 2018. Disponível em: <<https://www.bcb.gov.br/pec/wps/port/TD488.pdf>>. Acesso em: 1 maio 2022.

BRASIL, Banco da Amazônia. **Carta Anual de Públicas e Governança Corporativa 2021 – Exercício 2021**. Belém, 15 maio 2022. Disponível em: <<https://www.bancoamazonia.com.br/index.php/component/edocman/carta-anual-de-politicas-publicas-2021/viewdocument/5033>>. Acesso em: 19 maio 2022.

BRASIL. Banco do Brasil. **Carta Anual de Políticas Públicas e Governança Corporativa 2021- Ano Base 2020**. Brasília, 2021. Disponível em: <<https://ri.bb.com.br/governanca-e-sustentabilidade/carta-anual-de-politicas-publicas-e-governanca-corporativa/>>. Acesso em: 06 jan. 2022.

BRASIL, Banco do Nordeste. **Carta Anual de Públicas e Governança Corporativa Exercício 2020**. Fortaleza, 15 jun. 2021. Disponível em: <<https://bnb.gov.br/web/guest/transparencia-e-prestacao-de-contas/carta-anual-de-politicas-publicas-e-governanca-corporativa>>. Acesso em: 19 maio 2022.

BRASIL, Banco Central do Brasil. **Resolução n. 2.099, de 17 ago. 1994**. BCB, 1994. Disponível em: <https://www.bcb.gov.br/pre/normativos/res/1994/pdf/res_2099_v21_L.pdf>. Acesso em: 12 maio 2022.

BRASIL, Banco Central do Brasil. **Resolução n. 3.490, de 29 ago. 2007**. BCB, 2007. Disponível em: <https://www.bcb.gov.br/pre/normativos/res/2007/pdf/res_3490_v3_P.pdf>. Acesso em: 12 maio 2022.

BRASIL, Banco Nacional do Desenvolvimento – BNDES. **Carta Anual de Políticas Públicas e Governança Corporativa 2021- Ano Base 2020**. Rio de Janeiro, 2021. Disponível em: <<https://www.bndes.gov.br/wps/portal/site/home/transparencia/prestacao-de-contas/cartas-anuais-de-politicas-publicas-e-governanca-corporativa-do-bndes>>. Acesso em: 10 jan. 2022.

BRASIL, Caixa Econômica Federal. **Relatório Integrado 2020**. Brasília, 2021. Disponível em: <https://www.caixa.gov.br/Downloads/caixa-governanca/Relatorio_Integrado_Caixa_2020.pdf>. Acesso em: 13 jan. 2022.

BRASIL, Conselho de Monitoramento e Avaliação de Políticas Públicas. **Relatório de Avaliação Fundos Constitucionais de Financiamento do Norte (FNO), Nordeste (FNE) e Centro-Oeste (FCO)**. Brasília, 2021. Disponível em: <<https://www.gov.br/economia/pt-br/aceso-a-informacao/participacao-social/conselhos-e-orgaos-colegiados/cmap/politicas/2020/subsidios/relatorio-de-avaliacao-cmas-2020-fcfs>>. Acesso em: 22 dez. 2021.

BRASIL, Lei n. 4.595, de 31 de dezembro de 1964. Dispõe sobre a Política e as Instituições Monetárias, Bancárias e Creditícias, Cria o Conselho Monetário Nacional e dá outras providências. **Diário Oficial da União**, página 12081 de 31 dez. 1964.

CAMPOS, M.C. O papel da governança corporativa dos bancos públicos federais nos investimentos em infraestrutura. São Paulo, **FGV**, 2021. Disponível em: <<https://bibliotecadigital.fgv.br/dspace/handle/10438/31889>>. Acesso em: 8 set. 2022.

CROCCO, M.; NOGUEIRA, M.; ANDRADE, C.M.C.; DINIZ, G.F.C. O Estudo do Gap Regional de Crédito e seus Determinantes, sob uma Ótica Pós-Keynesiana. **Revista Economia**, 2011. Disponível em:<https://anpec.org.br/revista/vol12/vol12n2p281_307.pdf>. Acesso em: 12 jan. 2022.

CUNHA, G. A. C. (2020). Déficit habitacional: o tamanho da desigualdade social no Brasil. **Boletim Economia Empírica**, 2020. Disponível em: <<https://www.portaldeperiodicos.idp.edu.br/bee/article/view/4014>>. Acesso em: 8 set. 2022.

FERREIRA, G. G.; CALMON, P.; FERNANDES, A. S. A.; ARAÚJO, S. M. V. G. Política habitacional no Brasil: uma análise das coalizões de defesa do Sistema Nacional de Habitação de Interesse Social versus o Programa Minha Casa, Minha Vida. urbe. **Revista Brasileira de Gestão Urbana**, 11, 2019. Disponível em: <<https://www.scielo.br/j/urbe/a/jmR7Y6NdPqRwZf3PgQsVNjN/?lang=pt&format=pdf>>. Acesso em: 8 set. 2022.

HENRIQUES, D.F. **Uma leitura sobre inflação alta e indexação contratual no Brasil: origens e consequências**. 78p. Universidade Federal Fluminense, Niterói, 2020.

HORN, C.H.; FEIL, F. Instituições financeiras de desenvolvimento regional e os desafios do Sistema Nacional de Fomento. **Economia e Sociedade**, Campinas, v. 28, n. 1. p.227-254, 2019. Disponível em <<https://www.scielo.br/j/ecos/a/jxLgxP5R8zSQnkDfq7mHmNy/?lang=pt&format=pdf>> . Acesso em: 17 mai. 2022.

LEITE, K.V.B.S.; REIS, M. O Acordo de Capitais de Basiléia III: Mais do Mesmo? **Revista Economia**, v.24, pág 159-187, 2013. Disponível em: < http://anpec.org.br/revista/vol14/vol14n1p159_187.pdf >. Acesso em: 11 mai. 2022.

LEITE, S.P.; WESZ JUNIOR, V.J. Estado, Políticas Públicas e Agronegócio no Brasil: Revisitando o papel do Crédito Rural. R. **Pós Ci. Soc.** v.11, n.22, jul/dez. 2014. Disponível em: < http://devrima.im.ufrj.br:8080/jspui/bitstream/1235813/5559/1/Estado%2c%20pol%3%adticas%20p%3%ablicas%20e%20agroneg%3%b3cio%20no%20Brasil%20_%20revisitando%20o%20papel%20do%20cr%3%a9dito%20rural.pdf>. Acesso em: 6 set. 2022.

LOBATO, F. T. A Importância dos Bancos Públicos Brasileiros na Crise Financeira de 2008: Uma Visão Pós-Keynesiana. **A Economia em Revista - AERE**, v. 28, n. 1, p. 79-98, 2020.

METTENHEIM, K. Vantagens Competitivas Institucionais de Bancos Públicos. Rio de Janeiro, UFRJ: **Revista Desenvolvimento em Debate**, v.3, n.2, p.7-27, 2015. Disponível em: <<https://revistas.ufrj.br/index.php/dd/article/download/31901/18063>>. Acesso em: 2 out. 2022.

MONTEIRO, A.R.; VERAS, A.T.R. A questão habitacional no Brasil. **Mercator**, Fortaleza, v. 16, 2017. Disponível em : < <https://www.scielo.br/j/mercator/a/ZkVrVHZqbHWQwK6HRpGrcXN/?lang=pt&format=pdf>>. Acesso em: 8 set. 2022.

QUAGLIO, G.M. O Aspecto Financeiro do Desenvolvimento Regional: Uma Abordagem Pós-Keynesiana. **Iniciativa Econômica**, v.1, n.1. 2014. Disponível em:< <https://periodicos.fclar.unesp.br/iniciativa/article/view/6625>>. Acesso em: 11 jan. 2022.

RAMOS, S.Y.; MARTHA JUNIOR, G.B. Evolução da Política de Crédito Brasileira. Planaltina, DF: **Embrapa Cerrados**, 2010. Disponível em: <<https://ainfo>>.

cnptia.embrapa.br/digital/bitstream/item/77790/1/doc-292.pdf>. Acesso em: 6 set. 2022.

RÊGO, E.C.L. As regras de Prudência Bancária do Acordo de Basiléia. **Revista do BNDES**, Rio de Janeiro, 1995. Disponível em: <https://web.bndes.gov.br/bib/jspui/bitstream/1408/11294/1/RB%2003%20As%20Regras%20de%20Prud%c3%aancia%20Banc%c3%a1ria%20do%20Acordo%20de%20Basil%c3%a9ia_P_BD.pdf>. Acesso em: 12 mai. 2022.

SOIHET, E.; CABRAL, C.M.N. **Insights on monetary policy and public banking in Brazil (2008-2012)**. Nova Economia, 2016. Disponível em: <<https://rockcontent.com/br/talent-blog/referencia-bibliografica-abnt/>>. Acesso em: 8 jan. 2022.

SILVA, S.W.; GONÇALVES, J.E.; SOUZA, D.V.R.; PEREIRA, W.F.; FONSECA, L.R. O SISTEMA FINANCEIRO NACIONAL BRASILEIRO: contexto, estrutura e evolução. **Revista Universidade Vale do Rio Verde**, 2016. Disponível em: <<file:///C:/Users/MONICAR.ADM/Documents/TCC/Artigos/3043-11006-1-PB.pdf>>. Acesso em: 8 set. 2022.

TUROLLA, F. A.; GABRIELLI, M. F.; GONDIM, I. J. C. Crédito e Financiamento à Infraestrutura no Brasil. Tecnologia de Crédito. **Serasa-Experian**, v. 1, p. 1, 2013. Disponível em: <https://pesquisa-eaesp.fgv.br/sites/gvpesquisa.fgv.br/files/arquivos/gabrielli_-_credito_e_financiamento.pdf>. Acesso em: 8 set. 2022.



**TEMA: DESMATAMENTO
ILEGAL DA AMAZÔNIA**

FERRAMENTA:
ESTATÍSTICA DESCRITIVA

ANÁLISE DA APLICAÇÃO DA SANÇÃO DO EMBARGO NO COMBATE AO DESMATAMENTO ILEGAL NA AMAZÔNIA¹

RAFAEL LEAL GARCIA⁽¹¹⁾

INTRODUÇÃO

“Defender e melhorar o meio ambiente para as atuais e futuras gerações se tornou uma meta fundamental para a humanidade.”²

Uma das grandes preocupações mundiais, especialmente nas últimas décadas, é o desenvolvimento sustentável, priorizando atividades que protejam e renovem os recursos ambientais. Nesse contexto, a Amazônia, maior floresta tropical do mundo, exerce papel de destaque, pois colabora para o equilíbrio ecológico em todo o planeta.

1 Este artigo resulta de trabalho de conclusão orientado pelo Professor Francisco Eduardo Soares Paiva.

2 Trecho da Declaração da Conferência da ONU sobre o Meio Ambiente (Estocolmo, 1972).

Desse modo, o desmatamento na Amazônia é um problema que chama a atenção mundial, pois, além de contribuir para o efeito estufa que causa as mudanças climáticas, gera outros impactos negativos para a sociedade e o meio ambiente, tais como ameaça a espécies da fauna e da flora com a destruição de habitats, comprometimento da oferta hídrica e alteração do fluxo das chuvas.

Nesse sentido, têm crescido as demandas globais por desmatamento zero, e vários países vêm adotando medidas para diminuir as suas contribuições indiretas ao desmatamento por meio do controle ou da proibição de importações de produtos agropecuários (CAIO, 2022; e SØNDERGAARD, 2021).

No Brasil, após um período, entre 2005 e 2012, com forte redução das taxas de desmatamento (83%), observa-se um aumento gradual do desmatamento na região Amazônica desde 2013.

Gráfico 1 | Taxa de desmatamento na Amazônia Legal



Fonte: Inpe

Para coibir o desmatamento, o Estado brasileiro possui alguns instrumentos, dentre os quais a fiscalização ambiental é o principal deles. No entanto, existem problemas para garantir a eficácia dos autos de infração lavrados. Há

lentidão no julgamento dos processos de autos de infração pelo Ibama (CGU, 2019) e baixa arrecadação das multas ambientais (IBAMA, 2019). Configura-se, assim, baixa eficácia das multas aplicadas nos autos de infração, uma vez que não se tem, na prática, sanção imediata ao infrator.

Nesse sentido, a sanção administrativa do embargo se torna um relevante instrumento de combate ao desmatamento, pois, ao impor uma punição instantânea e ao atender à estratégia de descapitalização dos infratores, potencializa o efeito dissuasório da ação fiscalizatória (SCHMITT, 2015).

Assim, procurar compreender a dinâmica da aplicação do embargo foi a questão central que motivou a realização do estudo.

Nesse contexto, o objetivo do trabalho foi analisar como o Ibama, principal órgão federal responsável pela fiscalização ambiental, tem empregado o embargo de áreas e as atividades para combater o desmatamento ilegal na Amazônia e como essas ações se relacionam com as taxas de desmatamento.

Para tanto, buscou-se: (a) descrever a dinâmica e as características dos embargos aplicados pelo Ibama; (b) analisar a evolução dos embargos e dos desembargos realizados; (c) analisar a evolução dos embargos emitidos frente aos autos de infração; (d) analisar a evolução dos autos de infração lavrados por descumprimento do embargo; e (e) relacionar a aplicação dos embargos, dos desembargos e dos autos de infração lavrados com as taxas de desmatamento na Amazônia Legal.

Assim, espera-se que o resultado deste trabalho possa preencher lacunas de conhecimento sobre o tema, servir de insumo para o aperfeiçoamento das estratégias de fiscalização ambiental, visando ao combate ao desmatamento ilegal na Amazônia, bem como dar subsídios a trabalhos de auditoria mais específicos sobre a questão.

REFERENCIAL TEÓRICO

O trabalho foi pautado por estudos relacionados à efetividade da fiscalização ambiental do Ibama.

De acordo com Matos (2016), fazer a repressão baseada apenas na aplicação de multas não é a forma mais eficaz de combater o desmatamento ilegal. Deve-se procurar descapitalizar ou dificultar o acesso a programas de fomento do governo para as pessoas que praticam o desmatamento ilegal, a fim de asfixiá-las financeiramente.

Nessa linha, Schmitt (2015) destaca que a grande relevância da medida administrativa do embargo para o combate ao desmatamento na Amazônia é que a transferência de risco ao mercado potencializa o efeito de dissuasão, pois, além de gerar as consequências das sanções administrativas, faz com que o mercado restrinja oportunidades de negócio. Ao final, concluiu-se que a redução do desmatamento ilegal poderia ser maior se houvesse ampliação da quantidade de áreas embargadas.

Os estudos, portanto, convergem ao apontar que, por causar restrição econômica imediata ao infrator, o embargo de áreas desmatadas é uma das medidas mais eficazes de combate ao desmatamento.

Em relação ao cumprimento dos embargos, os estudos de Veríssimo (2022) e do XXVII Congresso Brasileiro de Cartografia (2018) concluíram, respectivamente, que em 86,9% e 70% das áreas embargadas analisadas houve atividade econômica, configurando expressivo descumprimento da medida do embargo.

METODOLOGIA

Considerando o escopo apresentado, a pesquisa utilizou abordagem quantitativa, com a análise dos dados baseada em técnicas de estatística descritiva. Além disso, foram criados alguns índices relacionando as variáveis de estudo, a fim de analisar a evolução do emprego do embargo.

Quanto à fonte, o trabalho utilizou os dados do Portal de Dados Abertos do Ibama e as taxas anuais de desmatamento na Amazônia medida pelo Inpe, no âmbito do Prodes³.

O foco de estudo da pesquisa foi a Amazônia Legal, que abrange municípios de nove estados: Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins, e teve como recorte temporal um período de mais de 11 anos, entre 1º/1/2012 e 22/8/2022⁴.

TERMO DE EMBARGO

No exercício do poder de polícia ambiental, além da multa ambiental pecuniária, há infrações que comportam a aplicação de medidas administrativas cautelares que acompanham o auto de infração, como os termos de embargo. Segundo a legislação ambiental, o embargo de obra ou a atividade e suas respectivas áreas têm por objetivo impedir a continuidade do dano ambiental, propiciar a regeneração do meio ambiente e dar viabilidade à recuperação da área degradada (BRASIL, 2008).

Para conseguir cessar a penalidade do embargo, isto é, desembargar a área e/ou a atividade, o autuado deverá apresentar documentação que comprove a sua regularidade. Uma vez embargada a área ou a atividade, o seu descumprimento enseja a aplicação cumulativa da sanção de suspensão da atividade.

3 Prodes: Projeto de Monitoramento da Floresta Amazônica Brasileira por Satélite

4 Sobre o recorte temporal, vale registrar que, inicialmente, pretendia-se utilizar os dados desde 2008, pois permitiria avaliar e comparar a evolução do comportamento dos embargos e dos autos de infração lavrados frente a períodos em que as taxas de desmatamento estavam tanto em tendência de queda (2008-2012), quanto em tendência de alta (2012-2021). No entanto, devido às limitações observadas nos dados dos termos de embargos, os quais se verificou que estão mais estruturados e com maior confiabilidade somente a partir de 2012, adotou-se esse como o ano inicial do período de estudo.

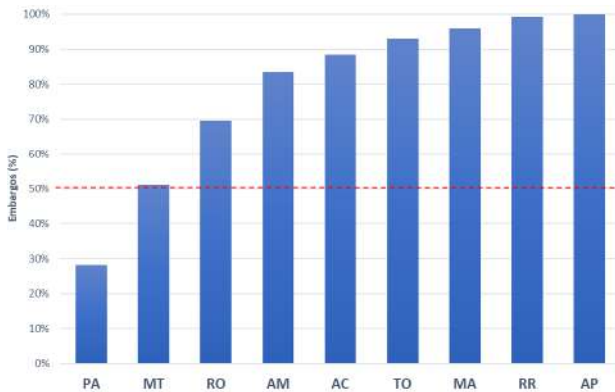
RESULTADOS ALCANÇADOS

Dinâmica e características dos embargos aplicados pelo Ibama

Distribuição espacial dos embargos

Na análise da distribuição dos embargos pela sua localização, verificou-se que os municípios que fazem parte da Amazônia Legal concentram 69% de todos os embargos aplicados pelo Ibama no período de 2012 a 2022. Dentro desse universo, foi possível notar que os Estados do Pará e do Mato Grosso possuem significativa relevância em termos quantitativos. Juntas, essas duas unidades da federação englobam cerca de 50% dos embargos na Amazônia Legal.

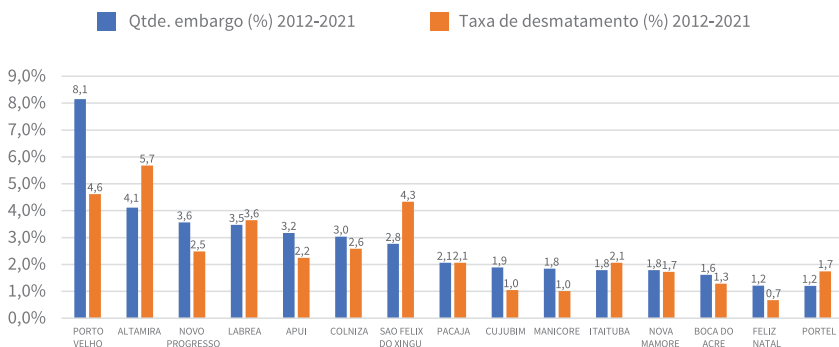
Gráfico 2 | Embargos na Amazônia Legal por UF



Fonte: Portal de Dados Abertos do Ibama (organizado pelo autor)

Descendo ao nível dos municípios, destaca-se que, dos 15 primeiros municípios com mais embargos, 12 encontram-se também entre aqueles que apresentam as maiores taxas de desmatamento. Em números totais, esses municípios representam 41% de todos os embargos aplicados pelo Ibama e 37% do total acumulado de área desmatada desde 2012:

Gráfico 3 | Principais municípios da Amazônia Legal com embargos e respectivas taxas de desmatamento (2012-2021)



Fonte: Portal de Dados Abertos do Ibama e Inpe (organizado pelo autor)

Assim, vê-se que há correspondência entre as regiões onde se efetuam os embargos e onde ocorrem o desmatamento, indicando que o Ibama tem atuado nas áreas mais críticas de desmatamento.

Distribuição dos embargos por natureza

A natureza do embargo representa o tipo de atividade que motivou o embargo.

Analisando-se a distribuição dos embargos na Amazônia Legal sob esse prisma, verifica-se que 65% deles são decorrentes de atividades de desmatamento. A título comparativo, quando se avalia essa distribuição fora da zona da Amazônia Legal, percebe-se que os embargos advêm de atividades mais diversas, especialmente em cinco atividades: desmatamento (29%), atividade de criador (19%), poluição/degradação (16%), atividade de exploração (10%) e empreendimentos (8%).

Distribuição dos embargos por item embargado

O item de embargo, de acordo com a classificação do Ibama, refere-se ao tipo de objeto embargado, como fazenda, floresta nativa, construção civil, mata nativa de preservação permanente etc.

Na Amazônia Legal, 69% dos embargos, conforme esperado, tiveram como objeto a floresta nativa, seguido por fazendas, com 9%; e madeiras, com 2%. Já fora da Amazônia, não se verifica alta concentração em um único objeto, conforme visto anteriormente, sendo criador amadorista de passeriformes (11,5%) e transporte de produtos perigosos (8,5%) os dois principais em termos quantitativos.

Distribuição dos embargos por tema ambiental da infração

Analisando a distribuição dos embargos em função do tema ambiental da infração, percebe-se que as infrações relacionadas à flora, como o desmatamento, representam quase que a totalidade dos autos de infração lavrados que originaram os embargos, com 88%.

Valor da multa dos autos de infração que possuem embargo associado

Na Amazônia Legal, o valor médio da multa por AI que possui embargo foi de R\$ 540.245,99. Já ao se analisar outra medida de tendência utilizada na estatística descritiva, a mediana, obtém-se o valor central de R\$ 155 mil para as multas, significando que metade desses autos de infração possuem multas de até esse valor.

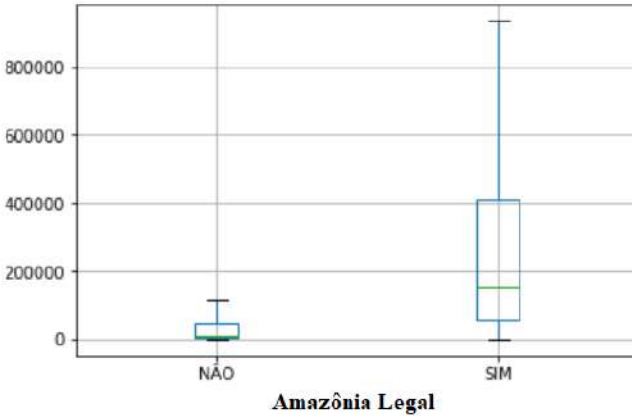
Essa expressiva diferença entre as duas medidas decorre do fato de que a média é mais suscetível à distorção por valores extremos, como são os valores das multas das infrações ambientais. Isso posto, avalia-se que, no caso em estudo, a mediana e os quartis conseguem representar melhor as faixas de valores em que se concentram os AI lavrados pelo Ibama⁵.

Assim, apurou-se que, na Amazônia Legal, 50% dos valores das multas aplicadas com embargos concentram-se entre R\$ 60 mil e R\$ 411.500. Quando se avaliam as multas fora da Amazônia, os valores caem para uma faixa entre R\$ 5 mil e R\$ 50 mil.

⁵ Para ilustrar os valores, utilizou-se o intervalo interquartil, que é uma medida estatística que avalia o grau de dispersão dos dados em torno da medida central. Trata-se da diferença entre o primeiro quartil (25%) e o quarto quartil (75%).

A seguir, apresenta-se um gráfico do tipo “*boxplot*” que ilustra essa diferença entre os valores dos AI com embargos dentro e fora da Amazônia Legal.

Gráfico 4 | Distribuição do valor da multa dos autos de infração com embargo



Fonte: Portal de Dados Abertos do Ibama (organizado pelo autor)

Diante das informações acima, fica visível que as multas dos autos de infração associados aos embargos no bioma amazônico são bastante superiores às multas lavradas em outras regiões. Nesse sentido, destaca-se que o valor central da multa na Amazônia (R\$155 mil) é quase 15 vezes superior ao valor encontrado para os AI fora desse bioma (R\$10.500).

Tal situação evidencia que os ilícitos na região amazônica possuem maior grau de lesividade ao meio ambiente, uma vez que o valor da multa está associado à dimensão da infração.

Tempo para desembargar uma área ou atividade

Outra variável estudada refere-se ao tempo que uma área ou uma atividade embargada levou para ser desembargada. Como resultado, apurou-se que, na Amazônia Legal, os autuados que conseguem realizar o desembargo demoram, em média, cerca de três anos para tanto. Utilizando novamente o

intervalo interquartil, tem-se que metade dos desembargos ocorrem entre 7,1 meses e 60,5 meses (5 anos).

Comparando-se esse tempo com os desembargos aplicados fora da Amazônia, não se percebe, diferentemente do observado no valor dos autos de infração, discrepância significativa.

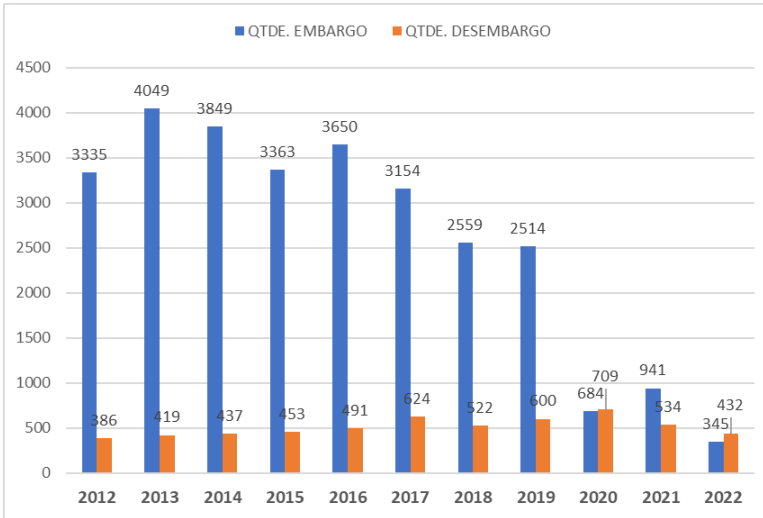
Embargos vigentes

Atualmente, 91% de todos os embargos aplicados a partir de 2012 estão vigentes, isto é, as áreas e as atividades continuam embargadas (não há registro de desembargo).

Portanto, consolidando as informações obtidas sobre as características e a dinâmica dos embargos aplicados pelo Ibama, verificou-se que essa sanção administrativa ocorre em maior grau na Amazônia Legal, principalmente no PA e MT, e são decorrentes, em sua maioria, de infrações de desmatamento ambiental com alto grau de lesividade ao meio ambiente, com valores da multa associada variando entre 60 e 400 mil reais. Além disso, verificou-se que grande parte dos embargos nessa região são mantidos e que os municípios com maiores taxas de desmatamento são também aqueles que concentram os embargos, indicando que o Ibama tem priorizado suas ações de fiscalização nessas regiões que sofrem maior pressão de desmatamento.

Análise da evolução dos embargos e desembargos realizados

No período de 1º/1/2012 a 22/8/2022, foram aplicados 28.443 embargos na Amazônia Legal, enquanto, no mesmo período, foram realizados 5.607 desembargos.

Gráfico 5 | Quantidade de embargos x desembargos

Fonte: Portal de Dados Abertos do Ibama (organizado pelo autor)

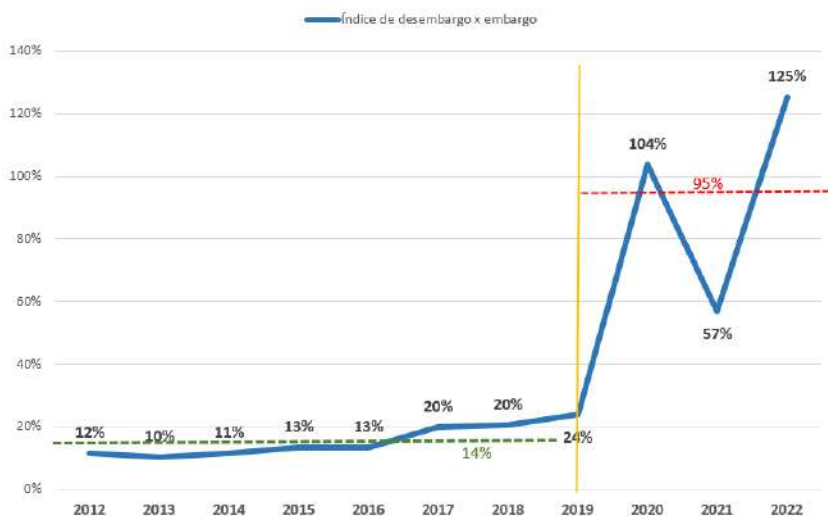
Analisando-se a série histórica, chama atenção a expressiva queda no número dos embargos, especialmente a partir de 2020. Ao mesmo tempo, a quantidade de desembargos tem apresentado leve tendência de alta.

Observa-se ainda que, nos anos de 2020 e 2022, a quantidade de desembargos foi superior ao de embargos. Nesses anos, o saldo anual de embargos (n° de embargos – n° de desembargos) foi negativo em 25 e 87 embargos, respectivamente. Essa situação ainda não havia ocorrido no período de estudo e, considerando que não foi um caso isolado, pode indicar nova tendência.

Para analisar a evolução entre a quantidade de embargos e desembargos, foi criado um índice que representa essa relação quantitativa:

$$\text{Índice de desembargo por embargo} = \left(\frac{\text{n. de desembargo}}{\text{n. de embargo}} \right) \times 100$$

O gráfico a seguir apresenta a evolução desse indicador.

Gráfico 6 | Índice de desembargo por embargo

Fonte: Portal de Dados Abertos do Ibama (organizado pelo autor)

Verifica-se que, até 2019, a média de desembargos por embargo aplicado pelo Ibama na Amazônia Legal era de 14%. Isto é, a cada 10 embargos realizados, aproximadamente 1 (um) desembargo era realizado. No entanto, nos três anos seguintes a 2019 — 2020, 2021 e 2022 — essa média disparou para 95%, o que representa um aumento de mais de 6 vezes a média histórica dos 8 anos anteriores.

Nesse ponto, vale destacar os números relativos ao ano de 2022 (com dados até o mês de agosto de 2022), em que a quantidade de desembargos aplicados chegou a 125% do número de embargos realizados, ou seja, para cada 10 embargos, cerca de 12 desembargos são feitos.

Vale registrar que os desembargos não necessariamente se referem aos mesmos embargos aplicados naquele ano. Na verdade, o mais comum são os desembargos serem feitos a partir de embargos aplicados em anos anteriores, conforme demonstrado no item referente ao tempo médio para desembargo.

Outro ponto que cabe destacar aqui é que o desembargo, por si só, não representa algo errado ou prejudicial ao meio ambiente. Como se trata de

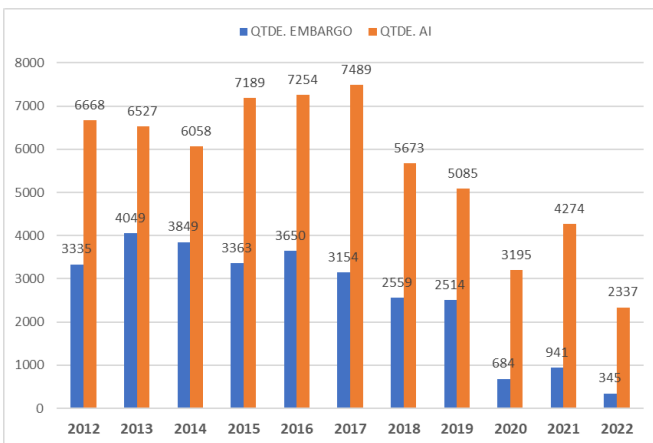
medida cautelar, no momento em que o autuado consegue comprovar, por meio de documentação, que seu empreendimento ou área que havia sido embargada encontra-se regular, o correto é o órgão realizar o desembargo.

No entanto, o que chama atenção não é somente o fato de o número de desembargos ter crescido, mas sim de essa situação ser acompanhada de substancial redução na aplicação de embargos pelo Ibama. Comparando-se os períodos de 2012 a 2019, com média de 3.309 embargos, com o período de 2020 e 2021, média de 813, verifica-se diminuição de 75% dos embargos lavrados.

Análise dos embargos emitidos frente aos autos de infração lavrados

No período de 1º/1/2012 a 22/8/2022, além dos 28.443 embargos, foram lavrados 61.749 autos de infração na Amazônia Legal. A evolução anual desses quantitativos é mostrada a seguir.

Gráfico 7 | Quantidade de embargos e de autos de infração lavado



Fonte: Portal de Dados Abertos do Ibama (organizado pelo autor)

De acordo com o gráfico, nota-se uma tendência de queda tanto na quantidade de autos de infração lavrados — chegando em 2020 ao menor nível histórico (anos completos) — quanto na de embargos aplicados. Uma vez que

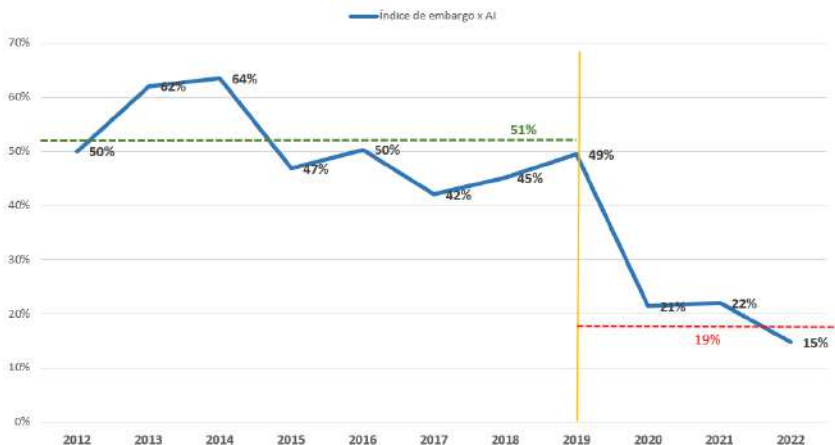
a aplicação dos embargos decorre de ação fiscalizatória com a lavratura de auto de infração, é natural que uma redução no número de autos provoque uma redução no número de embargos. Entretanto, objetivou-se verificar se essa redução ocorreu de forma proporcional.

Para tanto, foi criado um índice que representa a relação entre a quantidade de embargos aplicados frente a de autos de infração lavrados pelo Ibama, conforme a fórmula seguir:

$$\text{Índice de embargo por AI} = (\text{n. de embargos} / \text{n. de AIs}) \times 100$$

O gráfico abaixo apresenta a evolução do indicador.

Gráfico 8 | Índice de embargo por auto de infração



Fonte: Portal de Dados Abertos do Ibama (organizado pelo autor)

Na análise da evolução histórica, observa-se que, até 2019, a média de embargos por AI aplicado pelo Ibama na Amazônia Legal era de 51,4%, ou seja, metade dos autos de infração lavrados continham também a aplicação da medida cautelar do embargo. Destaca-se que nos anos de 2013 e 2014 o Instituto chegou a aplicar a medida do embargo em mais de 60% dos AI lavrados.

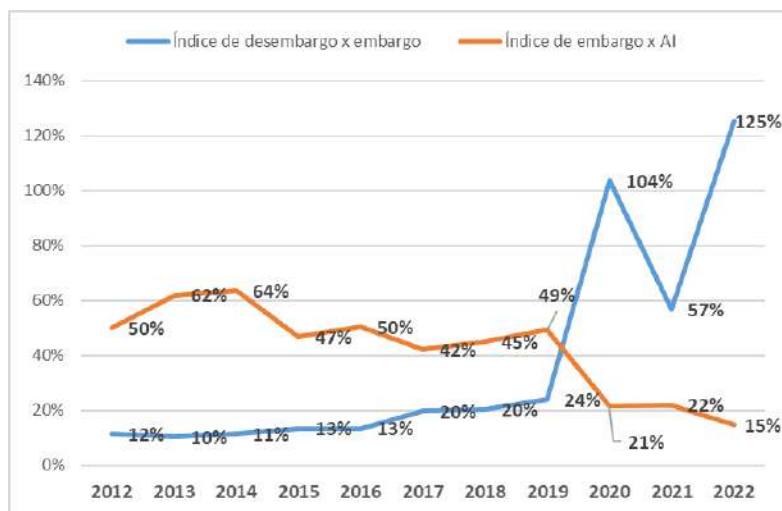
Ocorre que, nos anos posteriores a 2019 (2020, 2021 e 2022) essa média caiu para menos de 20%, isto é, menos de 2 (dois) embargos para cada 10

autos lavrados, sendo que em 2022 obteve-se o menor valor histórico para esse indicador, apenas 15%.

Essa significativa redução observada na aplicação da medida do embargo, proporcionalmente à lavratura dos autos de infração, pode indicar mudança na estratégia de fiscalização do Ibama nos últimos anos. Pode-se estar optando por atuar em áreas e/ou em tipos de infração que não permitam lavar termos de embargo conjuntamente aos autos de infração nas proporções anteriores, ou pode-se, por alguma razão, estar evitando aplicar a medida do embargo nos níveis anteriores, entre outras razões.

Adicionalmente, é importante frisar, conforme já visto, que o n. de desembargos frente ao de embargos vem seguindo o caminho oposto, com substancial ascensão:

Gráfico 9 | Índices de desembargo por embargo e de embargo por auto de infração



Fonte: Portal de Dados Abertos do Ibama (organizado pelo autor)

Portanto, dado que a aplicação do embargo se configura, conforme apresentado nos referencias teóricos aqui expostos, como uma das medidas mais eficazes de combate ao desmatamento, por causar restrição econômica

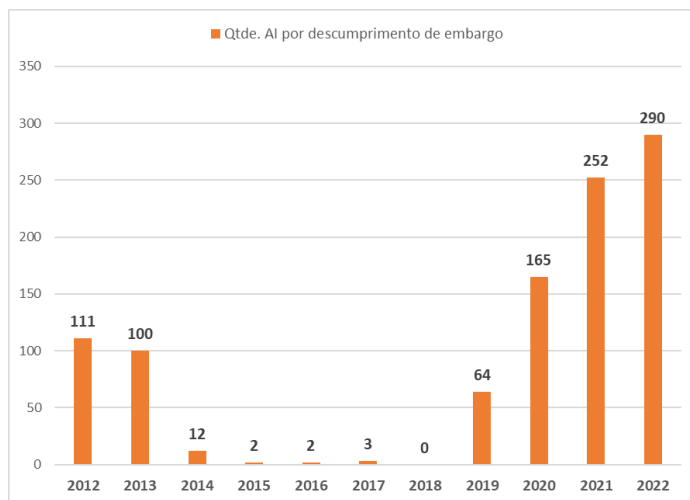
imediatamente ao infrator, as causas para essa mudança, evidenciadas pelos índices de embargo por AI e de desembargo por embargo, merecem ser melhor investigadas e aprofundadas em trabalho específico.

Análise da evolução dos autos de infração lavrados por descumprimento do embargo

A fim de avaliar se as áreas embargadas na Amazônia Legal têm sido objeto de ações de fiscalização para verificar o seu cumprimento, foram levantados todos os autos de infração lavrados pelo Ibama que possuíam o enquadramento legal das infrações relacionados a descumprimento de embargo⁶.

No período de 1º/1/2012 a 22/8/2022, foram emitidos 1.001 AI com infrações relativas ao descumprimento de embargo na Amazônia.

Gráfico 10 | Quantidade de autos de infração por descumprimento de embargo



⁶ Os três artigos do Decreto nº 6.514/2008 que tratam desse tema e foram selecionados são: Art. 18, Decreto 6.514/2008 - O descumprimento total ou parcial de embargo, sem prejuízo do disposto no art. 79, ensejará a aplicação cumulativa das seguintes sanções [...] Art. 54, Decreto 6.514/2008 - Adquirir, intermediar, transportar ou comercializar produto ou subproduto de origem animal ou vegetal produzido sobre área objeto de embargo. Art. 79, Decreto 6.514/2008 - Descumprir embargo de obra ou atividade e suas respectivas áreas.

Fonte: Portal de Dados Abertos do Ibama (organizado pelo autor)

A partir desses dados, pode-se concluir que, na prática, nos quatro anos compreendidos entre 2015 e 2018, o Ibama deixou de autuar e/ou fiscalizar as áreas e as atividades que haviam sido embargadas. Nesses anos, o total de AI lavrados por descumprimento do embargo foi de apenas 7 (sete), representando 0,02% do total lavrado.

Por outro lado, nos últimos anos, embora o número de autos de infração tenha diminuído, como já visto, verificou-se expressivo aumento do número de lavratura de AI por descumprimento de embargo.

Somente nos últimos 3 anos (2020 a 2022), lavrou-se mais do que o dobro dos AI por descumprimento de embargo emitidos no período de 8 (oito) anos, compreendido entre 2012 e 2018 (481 autos contra 230).

Tal situação sugere que o Ibama, embora tenha diminuído a aplicação do embargo, tem aumentado o foco de atuação para monitorar as áreas embargadas, verificando o seu cumprimento.

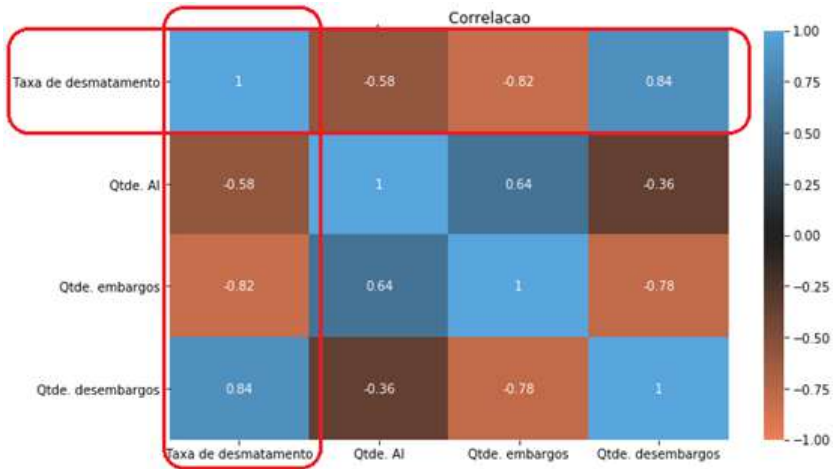
Análise da relação entre os embargos, desembargos e autos de infração lavrados com as taxas de desmatamento na Amazônia Legal

Além das análises quantitativas apresentadas, procurou-se avaliar como essas variáveis se relacionam com as taxas de desmatamento na Amazônia. Para isso, utilizou-se a correlação, que é uma medida que busca mensurar o grau de associação entre duas variáveis.

A partir dos dados de lavratura dos autos de infração, de emissão de embargos e desembargos realizados e da taxa de desmatamento na Amazônia Legal no período de 2012 a 2021, foi possível elaborar a seguinte matriz de correlação, com os respectivos coeficientes⁷:

⁷ Para interpretação dos resultados, vale registrar que o valor zero do coeficiente de correlação indica a inexistência de relação entre as duas variáveis e que quanto mais próximo de 1 for o valor absoluto desse coeficiente mais forte é a relação entre as duas variáveis. Os sinais positivo e negativo indicam o sentido da relação, respectivamente, direta ou inversa.

Gráfico 11 | Correlação entre a taxa de desmatamento e a quantidade de AI, embargos e desembargo



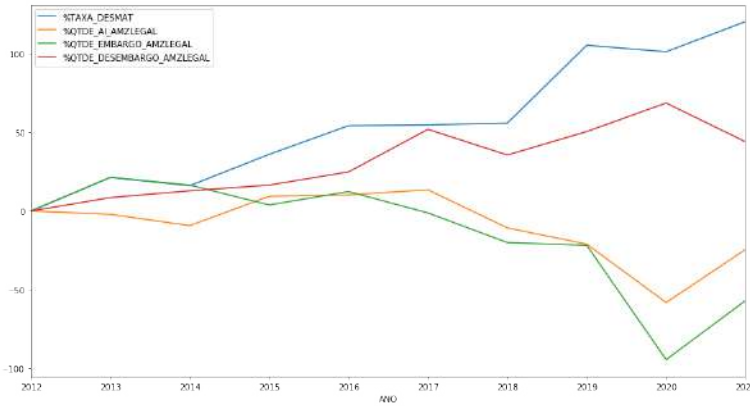
Fonte: Portal de Dados Abertos do Ibama e Inpe (organizado pelo autor)

Diante dos resultados obtidos, tem-se que a taxa de desmatamento demonstrou possuir: moderada correlação (inversa) com os autos de infração lavrados (-0,58); forte correlação (inversa) com a quantidade de embargos aplicados (-0,82); e forte correlação (direta) com o quantitativo de desembargos realizados (0,84)⁸. Ou seja, conforme os valores dos coeficiente de correlação indicam, a trajetória da taxa de desmatamento acompanha, no mesmo sentido, a de desembargos, enquanto a trajetória referente à quantidade de AI lavrados e à de embargos vai no sentido contrário ao do desmatamento.

Visualmente, tal fato pode ser ilustrado com a variação acumulada entre 2012 e 2021 de cada variável de estudo (embargos, desembargos, autos de infração e taxa de desmatamento), conforme a seguir:

⁸ Como referência para avaliar a intensidade da relação a partir do coeficiente, utilizou-se a tabela contida em Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. 5th ed. Boston: Houghton Mifflin; 2003.

Gráfico 12 | Evolução da variação acumulada da taxa de desmatamento, AI, embargos e desembargos



Fonte: Portal de Dados Abertos do Ibama e Inpe (organizado pelo autor)

Observa-se que enquanto a taxa de desmatamento aumentou 120% o número de AI lavrados caiu 24%, chegando no ano de 2020 a ter uma redução de 58%; a quantidade de embargos aplicados reduziu 57% no acumulado do período, chegando ao ápice, em 2020, com 94%; e a quantidade de desembargos, por outro lado, vem aumentando, tendo subido nesse período 44% em relação a 2012, sendo 2020 o ano de pico, tendo chegado a 68% de crescimento dos desembargos.

Assim, embora a existência de correlação não indique necessariamente uma relação causa-efeito entre as variáveis, no caso em particular das que foram objeto de estudo, é sabido que elas impactam diretamente nas taxas de desmatamento. Nesse sentido, vale trazer o que aponta o relatório de avaliação do Plano de ação para prevenção e controle do desmatamento na Amazônia Legal — PPCDam (3ª fase — 2012–2015):

Desde o ano de 2004, com a implementação do PPCDam, o monitoramento e o controle tem sido reconhecido em avaliações independentes como o que mais evoluiu e, conseqüentemente, como aquele com maior participação na redução de cerca de 77% na taxa de desmatamento entre 2004 (27.772 km²) e 2011 (6.418 km²).

A título de contextualização, cabe registrar que o eixo “monitoramento e controle” materializa-se, entre outras medidas, com as ações de fiscalização ambiental e os seus respectivos produtos, como autos de infração, termos de embargos, de apreensão de bens etc.

Portanto, dado o forte grau de correlação observado entre as variáveis de estudo com a taxa de desmatamento ambiental, pode-se concluir que a combinação da redução da fiscalização ambiental, materializada com as quedas nos autos de infração e nos embargos aplicados pelo Ibama, juntamente com o aumento dos desembargos, influenciou, além de outros fatores ocultos, no aumento das taxas de desmatamento na Amazônia Legal no período de 2012 a 2021.

CONCLUSÃO

A aplicação da medida do embargo de áreas e atividades, por causar restrição econômica instantânea ao infrator, é uma das medidas mais eficazes de combate ao desmatamento. Nesse contexto, o trabalho se propôs a analisar como o Ibama tem empregado o embargo de obras e atividades para combater o desmatamento ilegal na Amazônia e como essas ações se relacionam com as taxas de desmatamento.

Em relação aos embargos e aos desembargos, observou-se que, até 2019, somente 1 (um) desembargo era realizado para cada 10 embargos aplicados. No entanto, nos três anos seguintes — 2020, 2021 e 2022 — essa média disparou para 9 desembargos para cada 10 embargos. Chamou atenção não somente o fato de o número de desembargos ter crescido, mas sim dessa situação ser acompanhada de substancial redução de 75% dos embargos aplicados nos últimos três anos.

Verificou-se também significativa diminuição no uso da medida do embargo, proporcionalmente à lavratura dos autos de infração. Até 2019, metade dos autos de infração lavrados continham a aplicação da sanção do embargo.

Todavia, nos anos posteriores, essa média caiu para menos de dois embargos para cada 10 autos lavrados.

Essa mudança, evidenciada pelos números nos últimos anos, pode indicar alteração na estratégia de fiscalização do Ibama. Portanto, dado que a aplicação do embargo configura-se como uma das medidas mais eficazes de combate ao desmatamento, as causas para essa mudança merecem ser melhor investigadas.

Por fim, dado o forte grau de correlação observado entre as variáveis de estudo (embargos, desembargos e AI) com a taxa de desmatamento, pode-se concluir que a combinação da queda nos autos de infração e nos embargos aplicados pelo Ibama, juntamente com o aumento dos desembargos, influenciou, além de outros fatores ocultos, no aumento das taxas de desmatamento na Amazônia Legal no período de 2012 a 2021.

REFERÊNCIAS

BRASIL. Decreto n. 6.514, de 22 de julho de 2008.

CAIO, C. **Parlamento Europeu amplia proibição de commodities ligadas ao desmatamento**. Carta Capital, 13.09.2022. Disponível em: <<https://www.cartacapital.com.br/mundo/parlamento-europeu-amplia-proibicao-de-commodities-ligadas-ao-desmatamento/>>. Acesso em: 14 set. 2022.

CGU. Controladoria-Geral da União. **Relatório de Avaliação do Processo Sancionador Ambiental**. 2019.

IBAMA. Instituto Brasileiro do Meio Ambiente. **Relatório de Gestão do Ibama**. 2019.

MATOS, F. L. de C. C. **Análise das taxas anuais de desmatamento na Amazônia Legal a partir da relação entre autos de infração e área desmatada no período entre 2000 e 2014**. 90 f., il. Dissertação (Mestrado em Geografia) - Universidade de Brasília, Brasília, 2016.

MORAES, D. R. V. de. **Monitoramento de áreas embargadas por desmatamento ilegal**. Rev. Bras. de Cartografia, vol. 70, Edição Especial “XXVII Congresso Brasileiro de Cartografia”, 2018.

SCHMITT, J. **Crime sem castigo**: a efetividade da fiscalização ambiental para o controle do desmatamento ilegal na Amazônia. Tese (Doutorado em Desenvolvimento Sustentável). Universidade de Brasília, 2015.

SØNDERGAARD, N. **Demandas globais por desmatamento zero e a agropecuária brasileira**. Veja, 16.12.2021. Disponível em: <<https://veja.abril.com.br/coluna/agro-global/demandas-globais-por-desmatamento-zero-e-a-agricultura-brasileira/>>. Acesso em: 14 set. 2022.

VERISSIMO, C. S. da S. **Marked non-compliance with deforestation embargoes in the Brazilian Amazon**. Environmental Research Letters, 2022.

The background is a dark teal color with a complex network of glowing teal lines and nodes. The nodes are small circles of varying sizes, some of which are larger and more prominent. The lines connect these nodes in a web-like pattern, creating a sense of interconnectedness and data flow.

**TEMA: MEIO AMBIENTE
E PRODUÇÃO DE
BIOCOMBUSTÍVEL**

FERRAMENTA:
ESTATÍSTICA DESCRITIVA

PERFIL DOS MUNICÍPIOS PRODUTORES DE CANA-DE-AÇÚCAR EM RELAÇÃO A QUEIMADAS OU A INCÊNDIOS, À ÁREA COLHIDA E À PRESENÇA DE PRODUTOR DE BIOCOMBUSTÍVEL INCENTIVADO PELO PROGRAMA RENOVABIO¹

RODRIGO SANTOS DE PAULA⁽¹²⁾

INTRODUÇÃO

A redução da emissão de gases de efeito estufa (GEE²) é um objetivo mundial com metas firmadas pelos países que aderiram aos acordos estabelecidos no âmbito do tratado internacional da Convenção-Quadro das Nações Unidas

1 Este artigo resulta de trabalho de conclusão orientado pelo Professor Eduardo Soares Paiva.

2 GEE são substâncias gasosas que absorvem parte da radiação infravermelha, emitida principalmente pela superfície terrestre, e dificultam seu escape para o espaço. Isso impede que ocorra uma perda demasiada de calor para o espaço, mantendo a Terra aquecida. O efeito estufa é um fenômeno natural. Esse fenômeno acontece desde a formação da Terra e é necessário para a manutenção da vida no planeta, pois, sem ele, a temperatura média da Terra seria 33 °C mais baixa, impossibilitando a vida no planeta, tal como conhecemos hoje. O aumento dos gases estufa na atmosfera tem potencializado esse fenômeno natural, causando um aumento da temperatura (fenômeno denominado mudança climática). Disponível em: <https://pt.wikipedia.org/wiki/Gases_do_efeito_estufa>. Acesso em: 10 maio 2022.

sobre a Mudança do Clima (CQNUMC) ou ratificaram-nos, e seu mais recente compromisso advém da COP26, realizada em novembro de 2021, em Glasgow, na Escócia.

O Brasil, que ratificou o acordo de Paris negociado durante a COP21, recentemente atualizou seus compromissos estabelecidos por meio da Contribuição Nacionalmente Determinada (*Nationally Determined Contribution — NDC*) revista para o pacto de Glasgow.

Por meio desta comunicação, o Brasil confirma o compromisso apresentado em sua NDC revisada, submetida ao Secretariado da UNFCCC em 9 de dezembro de 2020, de reduzir as emissões de gases de efeito estufa em 37% abaixo dos níveis de referência de 2005, em 2025.

Adicionalmente, o país assume o compromisso de reduzir as emissões de gases de efeito estufa em 50% abaixo dos níveis de 2005, em 2030. ANDC revisada do Brasil antecipa para 2050, ainda, objetivo indicativo de longo prazo de alcançar a neutralidade climática. O escopo da NDC brasileira é amplo, prevendo-se a implementação de ações de mitigação e adaptação à mudança do clima ao conjunto da economia brasileira, assim como considerações sobre meios de implementação. (Brasil. Comunicação ao Secretariado da Convenção-Quadro das Nações Unidas sobre Mudança do Clima. 8 de fevereiro de 2022.)

Além da atualização de compromissos para redução da emissão de GEE, em sua última comunicação ao Secretariado da Convenção-Quadro das Nações Unidas sobre Mudança do Clima (UNFCCC), o Estado Brasileiro citou também a importância do programa RenovaBio para o crescimento da produção de biocombustíveis no país.

Cumpre lembrar, ainda, que o Brasil possui uma das matrizes energéticas mais limpas do mundo. Em 2020, fontes renováveis contabilizaram 48,4% da demanda total de energia, três vezes o indicador mundial. Na matriz de demanda total de energia elétrica o percentual de renováveis foi de 84,8%, e na matriz de consumo de combustíveis de transportes, foi de 25%. A produção de biocombustíveis para o setor de transportes tem crescido substancialmente por meio do programa RenovaBio, que utiliza incentivos de mercado para promover a descarbonização e aumentar a produção desse tipo de combustível oriundo de fontes renováveis.

O programa RenovaBio consiste em uma remuneração adicional dos produtores de biocombustíveis na proporção da eficiência energética atingida pelo seu processo produtivo em relação ao combustível fóssil equivalente. A ideia é recompensar as empresas produtoras na proporção em que o consumo do seu produto reduziria a emissão de GEE em comparação ao consumo equivalente em combustível fóssil.

A implementação dessa política pública se dá pela criação de um ativo denominado CBIO, cuja unidade equivale a “uma tonelada de carbono que deixa de ser emitida para a atmosfera”, segundo site da ANP — Agência Nacional do Petróleo, Gás Natural e Biocombustíveis —, principal agente operador do programa.

Cada produtor ou importador de biocombustível certificado no RenovaBio gera seus créditos em CBIO e os comercializam na bolsa de valores. O CBIO pode ser comercializado com quaisquer pessoas físicas ou jurídicas residentes ou não no Brasil, mas a principal demanda de compra desses ativos se dá por parte das empresas distribuidoras de combustíveis, que têm obrigações de aquisição de CBIOs em quantidades estabelecidas pela ANP conforme metas globais definidas pelo CNPE — Conselho Nacional de Política Energética.

O programa RenovaBio é recente, o primeiro certificado de produção foi aprovado em outubro de 2019 e outros 212 certificados foram aprovadas no

primeiro semestre de 2020, quando definitivamente o programa deslançou. Segundo dados do Caderno de Oferta de Biocombustíveis, referente aos Estudos do Plano Decenal de Expansão de Energia 2031 (PDE 2031), publicado em novembro de 2021 pela Empresa de Pesquisa Energética (EPE) e pelo Ministério de Minas e Energia (MME), a capacidade de produção em etanol certificada no RenovaBio correspondia a 87% da produção nacional. Considerando os dados de certificação divulgados pela ANP referentes à 16 de março de 2022, o número de certificações válidas estava em 307, sendo 261 relativas à produção de etanol de cana-de-açúcar de 1ª geração, largamente o biocombustível mais representativo na matriz energética brasileira.

O valor médio diário histórico de comercialização de um CBIO é de R\$ 50,04 (dados de 28/4/2020 a 20/4/2022), sendo que seu valor médio diário no primeiro trimestre de 2022 esteve em R\$ 85,97³. O volume de CBIO emitido está em crescimento, tendo sido registrados os seguintes números: 18.711.209 em 2020, 34.873.343 em 2021 e 17.365.030 somente no primeiro trimestre de 2022. Segundo site⁴ do MME, a negociação de CBIO em 2021 movimentou R\$ 1,17 bilhão.

Analisando-se os dados da Produção Agrícola Municipal 2020 — PAM 2020 —, constata-se que a cana-de-açúcar é a terceira cultura com maior área plantada no Brasil, ficando atrás somente do milho e da soja:

Quadro 1 | Principais culturas temporárias no Brasil

Posição	Cultura	Municípios em que está presente	Total de área plantada (ha)
1	Soja	2388	37.201.992
2	Milho	5104	18.351.075
3	Cana-de-açúcar	3307	10.014.198
4	Feijão	4328	2.686.870

Fonte: PAM 2020 (IBGE).

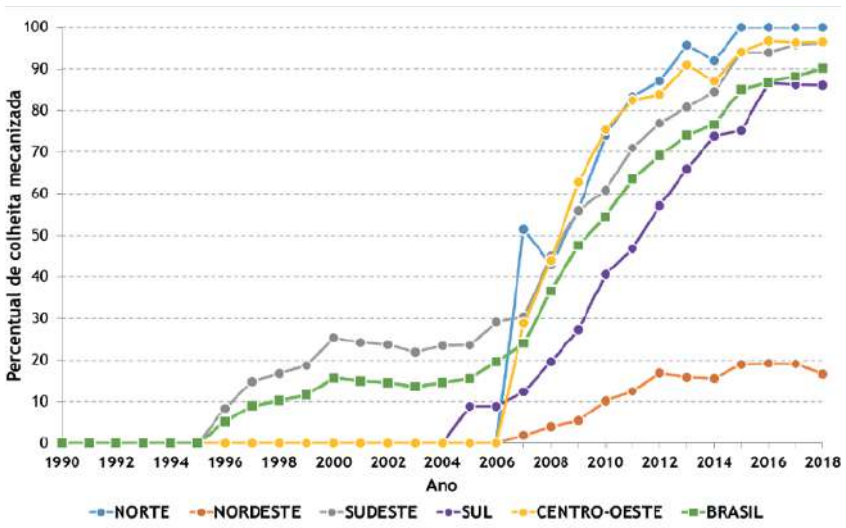
3 Perto do fechamento do artigo, verificou-se que o preço médio diário em junho de 2022 ficou em R\$ 162,94, tendo superado R\$ 200,00 em alguns dias. O preço se manteve ainda acima de R\$ 190,00 até meados de julho e depois caiu 50%, permanecendo oscilando entre R\$ 90,00 e R\$ 101,00 até o início de agosto de 2022.

4 Disponível em: <<https://www.gov.br/mme/pt-br/assuntos/noticias/mais-de-24-milhoes-de-toneladas-de-gases-de-efeito-estufa-foram-evitados-com-a-utilizacao-de-biocombustiveis-em-2021>>. Acesso em: 10 maio 2022.

A cultura da cana-de-açúcar vem se modernizado, mas uma prática comum é a realização da queima da palha previamente à colheita. Em função dos consequentes danos de ordem ambiental e de saúde pública, tal prática tem sido desestimulada ao longo dos anos, mas, dependendo da legislação local, ainda é permitida.

A partir de 2007, entretanto, intensificaram-se as medidas adotadas pela indústria sucroalcooleira no sentido de mecanizar o processo produtivo, o que substituiria a prática de queima da palha usualmente adotada antes da colheita para limpar o solo e viabilizar a colheita manual. Essa tendência de mecanização pode ser observada na Figura 1. Tais medidas teriam se iniciado especialmente a partir da implementação, pelo Estado de São Paulo (maior produtor), de protocolo, em acordo com as empresas produtoras, para eliminação dessa prática de queimada até o ano de 2017.

Figura 1 | Evolução da mecanização da colheita de cana-de-açúcar



Fonte: Con\$ab (2019).

Na Quarta Comunicação Nacional do Brasil à Convenção-Quadro das Nações Unidas sobre Mudança do Clima⁵ consta do item 2.5.6 – Queima de Resíduos Agrícolas (3.F), dado que vai ao encontro do efeito esperado pela mecanização da colheita:

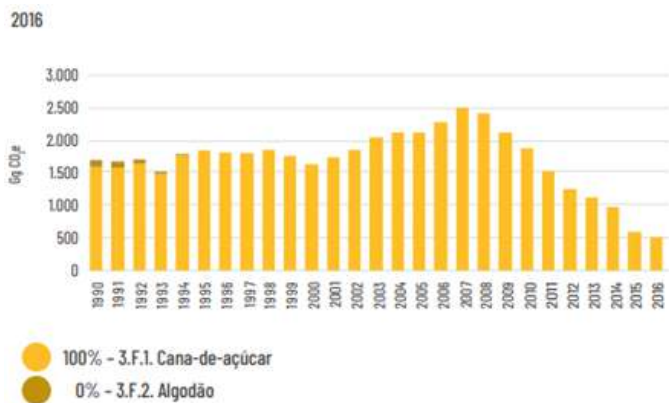
O subsetor Queima de Resíduos Agrícolas (3.F) contabilizou as emissões de CH₄ e N₂, decorrentes da queima realizada na pré-colheita da cana-de-açúcar e na pós-colheita de algodão herbáceo, sendo que este último ocorreu até 1994.

Em 2016, as emissões deste subsetor foram estimadas em 509 Gg CO₂ e de 2010 a 2016, observou-se uma redução de 72,8% nas emissões derivadas da queima de resíduos de cana-de-açúcar no país, apesar de o aumento na área colhida ter sido de 12,6%. Isso se deve ao processo de transição da colheita manual, que utiliza o fogo, para a colheita mecanizada (principalmente no estado de São Paulo), como se observa a partir de 2007, na Figura 2.31. As emissões de N₂O e CH₄ representaram 53% e 47% do total de CO₂ e do subsetor em 2016, respectivamente.

A Figura 2 a seguir ilustra a informação contida no referido documento.

5 Brasil. Ministério da Ciência, Tecnologia e Inovações. Secretaria de Pesquisa e Formação Científica. Quarta Comunicação Nacional do Brasil à Convenção Quadro das Nações Unidas sobre Mudança do Clima / Secretaria de Pesquisa e Formação Científica. — Brasília: Ministério da Ciência, Tecnologia e Inovações, 2021. ISBN: 978-65-87432-18-2

Figura 2 | Evolução das emissões decorrentes da queima de resíduos agrícolas (1990 a 2016)



Fonte: Quarta Comunicação Nacional do Brasil à Convenção Quadro das Nações Unidas sobre Mudança do Clima. (Adaptado)

Nos últimos 3 anos, foram registrados em território brasileiro pelo satélite de referência utilizado pelo INPE — Instituto Nacional de Pesquisas Espaciais (AQUA_M-T) — 604.510 focos de calor, sendo 197.632 em 2019, 222.797 em 2020 e 184.081 em 2021.

Conforme consta na seção Perguntas e Respostas do site do INPE⁶, Satélite de Referência é:

Mesmo indicando uma fração do número real de focos de queima e incêndios florestais, por usar o mesmo método de detecção e gerar imagens em horários próximos ao longo dos anos, os resultados do “satélite de referência” permitem analisar as tendências espaciais e temporais dos focos.

As imagens geradas pelo AQUA_M-T tem uma resolução de 1 km², ou seja, 1 pixel da imagem corresponde a 1 km², entretanto o INPE declara que testes de campo indicam que a sensibilidade da tecnologia utilizada no satélite consegue detectar frentes de fogo a partir de 30 m por 1 m de largura e, dependendo das condições, até menores.

6 Disponível em: <<https://queimadas.dgi.inpe.br/queimadas/portal/informacoes/perguntas-frequentes#p7>>. Acesso em: 10 maio 2022.

Nesse contexto, a pesquisa referente ao presente artigo buscou responder às seguintes perguntas:

- Com o processo de mecanização da colheita e a expressiva redução da emissão de gases de efeito estufa por queima de resíduos agrícolas identificada na pesquisa bibliográfica, os municípios produtores de cana-de-açúcar apresentam quantitativo de focos de calor equivalente aos demais ou há um perfil diferente?
- Considerando-se as três culturas agrícolas mais presentes no Brasil (soja, milho e cana-de-açúcar), os perfis dos municípios em relação à quantidade de focos de calor são semelhantes ou diferentes a depender da cultura predominante?
- Os municípios em que existe produtor de biocombustível certificado no programa RenovaBio têm área de cana-de-açúcar colhida e quantitativo de focos de calor semelhante ou diferente dos demais municípios produtores de cana-de-açúcar?

Para isso, utilizaram-se dados públicos e de fontes oficiais, conforme sintetizado a seguir:

- Identificação e dados geográficos dos municípios: IBGE.
- Mapas vetorizados dos municípios e dos biomas brasileiros: IBGE.
- Produção Agrícola Municipal — PAM 2020: IBGE.
- Produtores de etanol certificados no âmbito do programa RenovaBio: ANP (Dados de março/2022).
- Focos de Calor no território brasileiro: INPE (BDQueimadas). Dados de 2020 para as análises e de 2012 a 2021 somente para o cálculo da média, da mediana e do desvio padrão do histórico de 10 anos por bioma.

ESCOPO E PREMISSAS

- Focos de Calor: Todos os municípios brasileiros, em especial os que têm predominância⁷ dos biomas Mata Atlântica e Cerrado. Ano de 2020 para

⁷ Para identificação do bioma de cada município, foram aplicadas técnicas de análise vetorial no *software* QGIS de modo a estabelecer o bioma predominante para os municípios cujos territórios se estendam para mais de um bioma.

as análises territoriais e comparativas por cultura agrícola. Somente quantidade, sem avaliar duração, extensão ou origem (se queimada ou incêndio). Não está no escopo também a avaliação sobre a legalidade ou não de eventual queimada.

- Produção Agrícola Municipal: Dados sobre área plantada, por cultura, em todos os municípios brasileiros. Ano de 2020. Culturas: cana-de-açúcar, soja e milho (essas duas últimas para comparativos).
- Produtores Certificados RenovaBio: Todos os produtores em território brasileiro de etanol de primeira geração a partir da cana-de-açúcar certificados no programa.

Para identificação do município em que se produz etanol biocombustível de primeira geração a partir da cana-de-açúcar em que se tenha produtor certificado pelo RenovaBio, foi considerado o município indicado na planilha de certificados atualizada periodicamente pela ANP em seu site⁸. O município informado consta do campo “Emissor Primário”, no qual é identificada a usina. Para efeito de simplificação e considerando que não há dados estruturados públicos sobre as propriedades rurais que fornecem matéria-prima para as usinas, considera-se somente o município informado na planilha como um município produtor de cana-de-açúcar para etanol certificado no RenovaBio, não obstante a possibilidade de a usina também processar matéria-prima de outro município.

Quanto à identificação dos focos de calor, importa esclarecer que o satélite de referência utilizado pelo INPE registra as imagens no início da tarde e tem resolução de 1 km x 1 km para cada pixel. Testes de campo indicam que o satélite consegue detectar frentes de fogo com 30 m por 1 m de largura ou até menores.

8 Disponível em: <<https://www.gov.br/anp/pt-br/assuntos/renovabio/certificados-producao-importacao-eficiente-biocombustiveis>> .

DESENVOLVIMENTO

Antes de passar às análises, importa registrar que os focos de calor detectados pelos satélites podem ter origem em causas naturais, em intervenção humana ou em ambas. Nesse sentido, o presente projeto de pesquisa buscou avaliar um aspecto específico da intervenção humana, o cultivo da cana-de-açúcar. Considerando que o fenômeno analisado apresenta diversas variáveis que o influenciam diretamente e indiretamente, como índice pluviométrico, umidade do ar, atividade econômica desenvolvida, entre outros, buscou-se efetuar recortes e comparações que neutralizassem, em certo, grau essas variáveis. Pode-se citar como exemplos a utilização de recortes por biomas e mesorregiões e a comparação com o cultivo de outra cultura.

Pergunta 1: com o processo de mecanização da colheita e a expressiva redução da emissão de gases de efeito estufa por queima de resíduos agrícolas identificada na pesquisa bibliográfica, os municípios produtores de cana-de-açúcar apresentam quantitativo de focos de calor equivalente aos demais ou há um perfil diferente?

Para responder a essa pergunta serão realizadas análises focando nos municípios dos estados de São Paulo e de Goiás, que serão analisados separadamente. Os dois estados foram escolhidos por estarem entre os maiores produtores de cana-de-açúcar e por possuírem características diferentes, em especial o bioma.

Para ilustrar o perfil de focos de calor dos municípios, esses foram divididos em três grupos cuja regra de criação representa a presença da cana-de-açúcar em seu território. Sendo assim, os grupos foram definidos da seguinte forma: nenhuma ou baixa (municípios com baixa porcentagem de ocupação do território com a cultura de cana-de-açúcar), média (municípios com média ocupação do território com a cultura de cana-de-açúcar) e alta (municípios com alta taxa de ocupação de cana-de-açúcar em relação ao território). Cabe ressaltar que os índices de ocupação das áreas do município que foram considerados para enquadrá-los nas diferentes categorias diverge de acordo

com o estado considerado, conforme pode ser observado nos quadros 2 e 3. Essa estratégia foi utilizada pelo fato de os municípios dos dois estados analisados terem características muito diferentes, o que inviabilizaria uma análise utilizando-se critério único.

A análise focará no cruzamento de dois dados principais: o índice de focos de calor por 100 km² (ind_FC_100 km²) e o índice de área de cana-de-açúcar colhida por área territorial (indCana). O indCana é calculado para cada município individualmente e o ind_FC_100 km² é definido considerando a área total de cada grupo no estado.

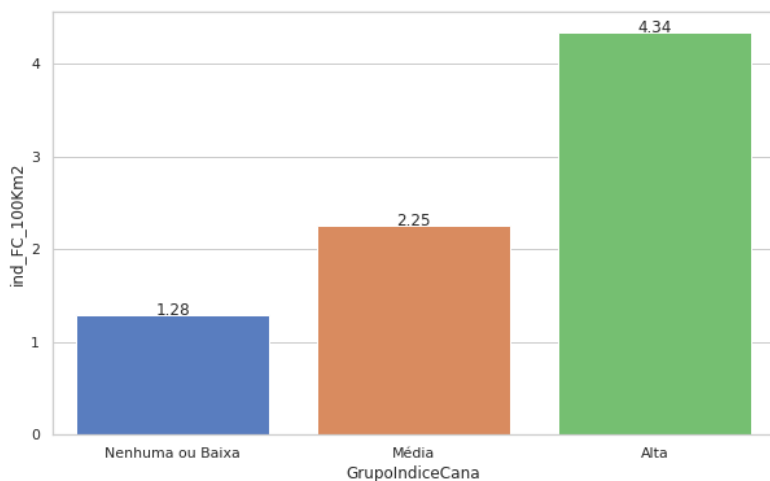
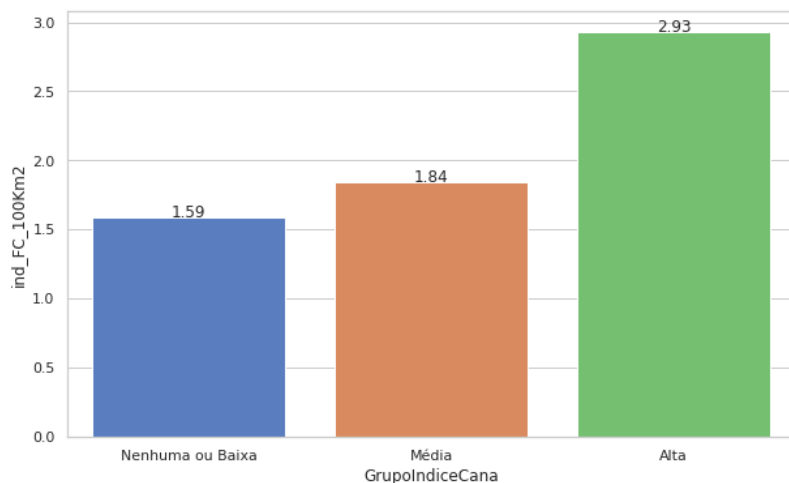
Os municípios da mesorregião metropolitana de São Paulo foram excluídos da análise, em função de ser uma área altamente conurbada.

Quadro 2 | Faixas adotadas para análise do estado de São Paulo

Distribuição das faixas em São Paulo		Equivalente a aproximadamente:
Nenhuma ou Baixa	indCana ≤ 1%	25% dos municípios, 1º quartil
Média	1% < indCana ≤ 41%	50% dos municípios, 2º e 3º quartis.
Alta	indCana > 41%	25% dos municípios, acima do 3º quartil.

Quadro 3 | Faixas adotadas para análise do estado de Goiás

Distribuição das faixas em Goiás		Equivalente a aproximadamente:
Nenhuma ou Baixa	indCana ≤ 1%	71% dos municípios.
Média	1% < indCana ≤ 10%	17% dos municípios.
Alta	indCana > 10%	12% dos municípios.

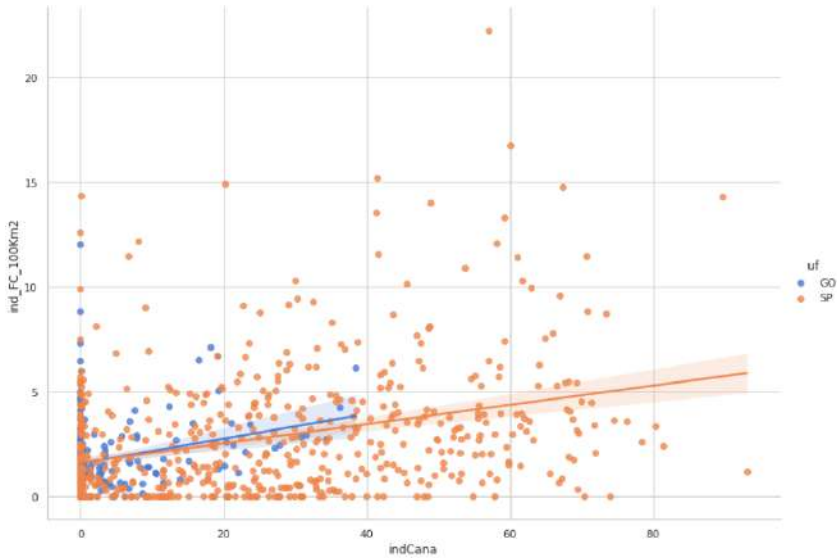
Figura 3 | Resultados de São Paulo conforme grupos estabelecidos**Figura 4 |** Resultados de Goiás conforme grupos estabelecidos

Para explorar os dados sem agrupamentos e comparar⁹ a posição dos municípios de ambos os estados considerando as duas variáveis analisadas, utilizou-se o gráfico da Figura 5. Esse gráfico apresenta a tendência de comportamento dos dados comparando a proporção de área de cana-de-açúcar

⁹ Apesar de os gráficos de barra mostrarem que o estado de São Paulo teria os grupos médio e alto com índices superiores aos de Goiás, esses dois grupos não devem ser comparados com os de mesmo nome do outro estado por terem sido estabelecidos com limites diferentes de indCana, conforme explicado.

colhida em relação ao território do município (indCana - eixo horizontal) com o número de focos de calor por 100 km² (ind_FC_100 km² - eixo vertical).

Figura 5 | Relação entre o Índice de Cana Colhida e o Índice de Focos de Calor por Município (Goiás e São Paulo)



Resultados específicos para a Pergunta 1:

- Em ambos os estados (SP e GO) foi identificada uma tendência de quanto maior o índice de área de cana colhida nos municípios, segundo dados da PAM-2020, maior é o índice de focos de calor.
- A amplitude da diferença entre os grupos Alto e Baixo em SP é maior que em GO, o que corrobora a diferença entre os desvios-padrão do ind_FC_100 km² nos estados, respectivamente, 22,33 e 7,24.
- Os municípios de Goiás com pouca ou nenhuma presença de área de cana-de-açúcar colhida em 2020 (barras azuis das Figuras 3 e 4) apresentaram ligeiramente mais focos de calor por 100 km² do que os municípios de São Paulo do mesmo grupo, o que é esperado, considerando a predominância do bioma Cerrado e o clima característico da região.

- É possível identificar na Figura 8 que a tendência de crescimento do ind_FC_100 Km^2 em função do indCana é semelhante nos dois estados, sendo em Goiás ligeiramente maior.

Pergunta 2: Considerando-se as três culturas agrícolas mais presentes no Brasil (soja, milho e cana-de-açúcar), os perfis dos municípios em relação à quantidade de focos de calor são semelhantes ou diferentes a depender da cultura predominante?

Considerando o expressivo número de municípios brasileiros, inclusive que plantam essas culturas em escala muito reduzida, foi estabelecido um limite mínimo de 5 km² de área colhida da respectiva cultura em 2020 para que os municípios compusessem os conjuntos de dados dessa análise específica.

As Figuras 6, 7 e 8 indicam os perfis dos municípios com relação ao índice de focos de calor por 100 km² em função da proporção de área colhida para as culturas de cana-de-açúcar, soja e milho, respectivamente.

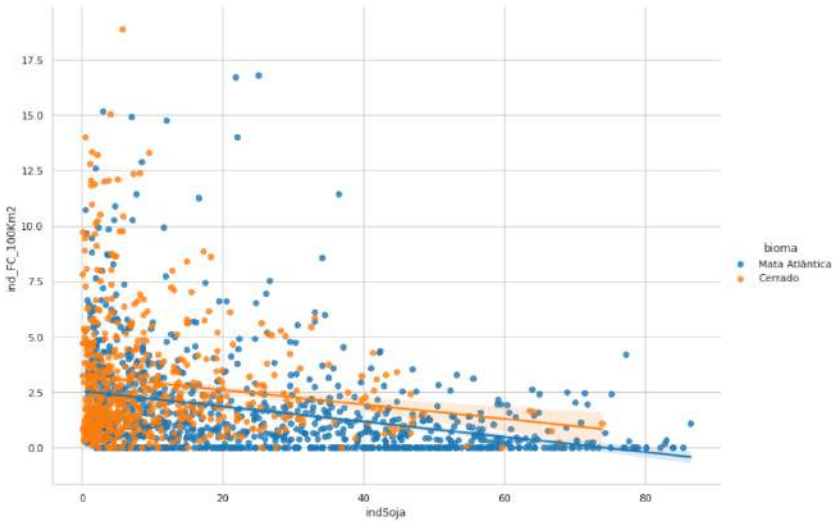
Cana-de-açúcar

Figura 6 | Relação entre o de Área de Cana Colhida e o Índice de Focos de Calor por Município (Mata Atlântica e Cerrado)



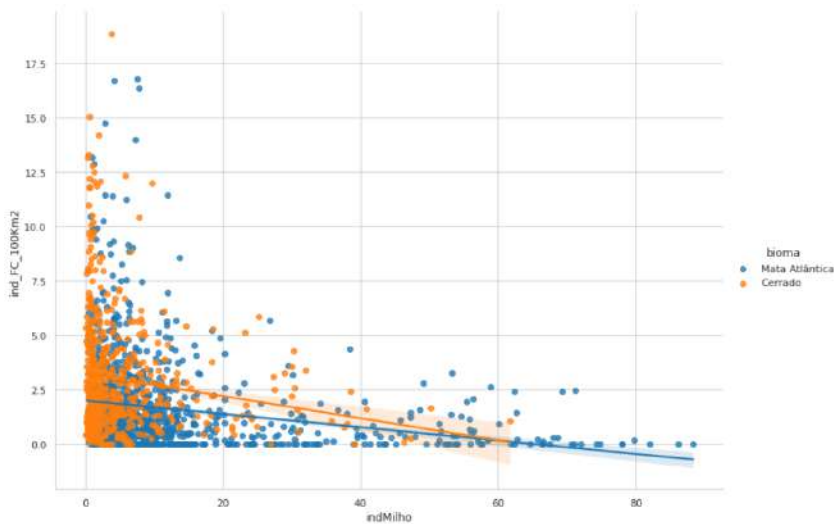
Soja

Figura 7 | Relação entre o Índice de Área de Soja Colhida e o Índice de Focos de Calor por Município (Mata Atlântica e Cerrado)



Milho

Figura 8 | Relação entre o Índice de Área de Milho Colhido e o Índice de Focos de Calor por Município (Mata Atlântica e Cerrado)



Constata-se, por meio da leitura da inclinação das retas das Figuras 7 e 8, que há uma tendência na redução da ocorrência de focos de calor nos municípios com cultivo de soja e milho. Dessa forma, quanto maior a área colhida de soja e de milho (em relação à área do município) em ambos os biomas analisados, Mata Atlântica e Cerrado, menor é a quantidade de focos de calor.

No entanto, uma tendência inversa é observada no caso da cana-de-açúcar. Sendo assim, a quantidade de focos de calor detectados tende a ser maior quanto maior for a área de cana colhida em relação à área territorial do município, também em ambos os biomas, conforme verifica-se na Figura 6.

Para analisar o dado de forma mais regionalizada e menos agregada, buscando-se, desse modo, reduzir a interferência de outras variáveis que influenciam o fenômeno analisado, foram selecionadas duas mesorregiões em que houvesse municípios tanto com predominância de cana-de-açúcar quanto de uma outra dessas culturas. Foram selecionadas as mesorregiões Norte Central Paranaense, cujo bioma predominante é a Mata Atlântica; e a Sul Goiano, onde o Cerrado predomina.

Cabe destacar inicialmente que um município, nas Figuras 10 e 12 a seguir, não é representado por um ponto exclusivamente, exceto se cada cor for analisada isoladamente, pois cada ponto corresponde ao índice da cultura respectiva no município (eixo x) associada ao índice de focos de calor (eixo y). Ou seja, se um município tem as três culturas presentes, ele aparecerá em três cores distintas, mas no mesmo nível do eixo y, pois o ind_FC_100 Km^2 é calculado por município, e não por propriedade rural produtora de uma ou de outra cultura.

Norte Central Paranaense

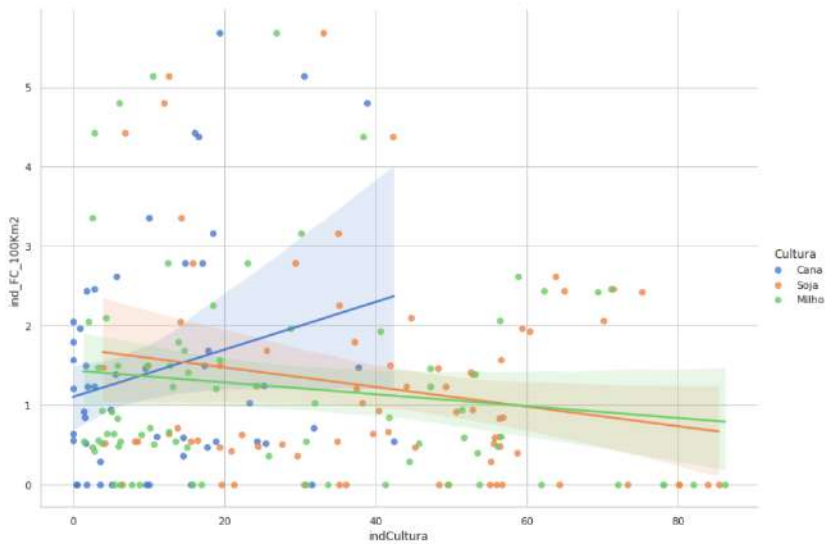
A mesorregião é composta de 79 municípios e faz fronteira com duas mesorregiões importantes produtoras do estado de São Paulo, Presidente Prudente e Assis.

Figura 9 | Mesorregião Norte Central Paranaense destacada no mapa do Estado



Fonte: Wikipedia. Acesso em: 10 jul. 2022.

Figura 10 | Relação entre o Índice de Área Colhida (Cana, Soja e Milho) e o Índice de Focos de Calor por Município do Norte Central Paranaense



Verificam-se na mesorregião as mesmas tendências anteriormente identificadas nacionalmente, redução na quantidade de focos de calor por 100 km² nos municípios com maior percentual de área colhida de soja e de milho e tendência de crescimento quanto maior o percentual de área colhida de cana-de-açúcar.

Para detalhar parte dos dados plotados na Figura 10, seguem, no Quadro 4, os 10 municípios da mesorregião que ocupam os extremos no ind_FC_100 Km² (5 maiores e 5 menores). Em caso de empate no extremo inferior, a área do município é usada como segundo fator de ordenação, priorizando exibir os de maiores áreas.

Quadro 4 | Extremos Norte Central Paranaense (Municípios com Maiores e Menores Índices de Focos de Calor)

5 municípios com maiores ind_FC_100km²					
Município	Bioma Predominante	indCana	indSoja	indMilho	ind_FC_100Km²
Florestópolis	Mata Atlântica	19,37	33	26,79	5,68
Presidente Castelo Branco	Mata Atlântica	30,51	12,68	10,59	5,14
Porecatu	Mata Atlântica	38,89	12	6,17	4,8
Miraselva	Mata Atlântica	16,14	6,81	2,77	4,43
Munhoz de Melo	Mata Atlântica	16,59	42,33	38,39	4,38
5 municípios com menores ind_FC_100 km²					
Município	Bioma Predominante	indCana	indSoja	indMilho	ind_FC_100Km²
Ibiporã	Mata Atlântica	0	55,95	49,54	0
Cruzmaltina	Mata Atlântica	0	35,22	15,85	0
São Jorge do Ivaí	Mata Atlântica	1,82	83,88	81,98	0
Rio Branco do Ivaí	Mata Atlântica	0	19,62	6,28	0
Marilândia do Sul	Mata Atlântica	0	56,71	8,71	0

O quadro de extremos corrobora a análise de tendência e verifica-se que, nos municípios com menor índice de focos de calor, a área de cana-de-açúcar colhida é quase inexistente, ao mesmo tempo da expressiva presença de soja

e de milho. No outro extremo, ou seja, nos municípios com maior índice de focos de calor, há considerável presença de cana-de-açúcar, mesmo nos dois (de cinco) municípios em que ela não foi a cultura predominante em 2020.

Sul Goiano

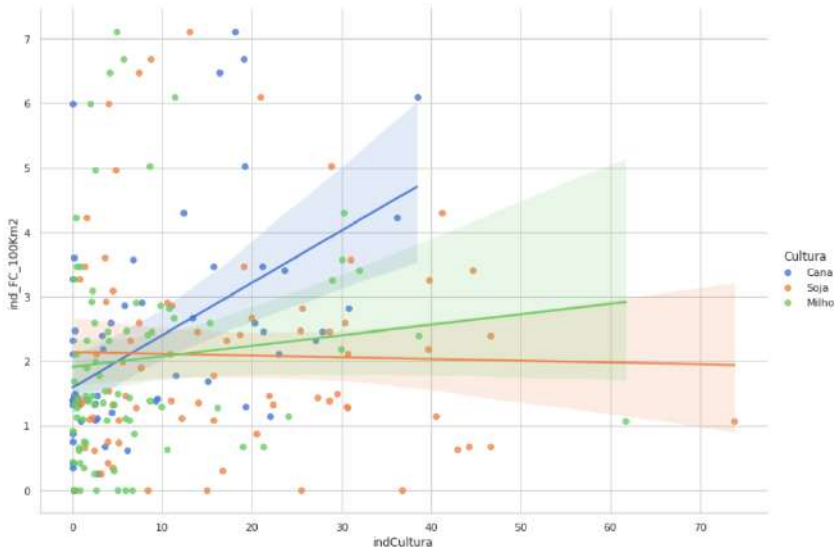
A mesorregião do Sul Goiano é a maior do estado de Goiás, sendo composta por 82 municípios e fazendo fronteira com 3 estados: Minas Gerais, Mato Grosso do Sul e Mato Grosso.

Figura 11 | Mesorregião Sul Goiano destacada no mapa do Estado



Fonte: Wikipedia. Acesso em: 10 jul. 2022.

Figura 12 | Relação entre o Índice de Área Colhida (Cana, Soja e Milho) e o Índice de Focos de Calor por Município do Sul Goiano



Do mesmo modo que nos gráficos anteriores, identifica-se na Figura 12 uma tendência de crescimento no índice de focos de calor à medida que se aumenta a proporção da área colhida de cana-de-açúcar em relação à área municipal, porém, nesse caso, a tendência identificada se mostrou mais acentuada.

Com relação à cultura de soja, observou-se tendência de neutralidade, enquanto, para a cultura de milho, foi identificada leve tendência de crescimento.

O comportamento do gráfico representado na Figura 12 é condizente com o bioma Cerrado, naturalmente mais suscetível à ocorrência de focos de calor. Percebe-se como que as linhas de tendências tivessem girado no sentido anti-horário e se iniciando em valores de ind_FC_100 km² ligeiramente mais altos do que na mesorregião Norte Central Paranaense, representada pela Figura 10.

Para detalhar parte dos dados plotados na Figura 12, seguem, no Quadro 5, os 10 municípios da mesorregião que ocupam os extremos no ind_FC_100 Km²

(5 maiores e 5 menores). Conforme já indicado, em caso de empate no extremo inferior, a área do município é usada como segundo fator de ordenação, priorizando exibir os de maiores áreas.

Quadro 5 | Extremos Sul Goiano (Municípios com Maiores e Menores Índices de Focos de Calor)

5 municípios com maiores ind_FC_100 km²					
Município	Bioma Predominante	indCana	indSoja	indMilho	ind_FC_100 Km²
Itumbiara	Cerrado	18,1	13,08	4,86	7,11
Castelândia	Cerrado	19,05	8,69	5,68	6,68
Panamá	Cerrado	16,43	7,4	4,16	6,48
Porteirão	Cerrado	38,43	20,95	11,38	6,1
Cezarina	Cerrado	0,02	4,08	1,99	5,99
5 municípios com menores ind_FC_100 km²					
Município	Bioma Predominante	indCana	indSoja	indMilho	ind_FC_100Km²
Aloândia	Cerrado	0	25,47	5,88	0
São João da Paraúna	Cerrado	0	14,98	2,58	0
Nova Aurora	Cerrado	0,13	0,29	0,81	0
Água Limpa	Mata Atlântica	0	0,21	0,15	0
Edealina	Cerrado	0,15	36,78	6,69	0

Da mesma forma que no item anterior, o quadro de extremos corrobora a análise de tendência. Nesse sentido, verifica-se que nos municípios com menores índices de foco de calor a área de cana-de-açúcar colhida é quase

inexistente. Dentre os cinco municípios com menos focos de calor, três tiveram consideráveis áreas de soja colhida e dois não apresentaram presença significativa de nenhuma das culturas analisadas.

Por outro lado, dentre os cinco municípios com maiores índices de focos de calor, quatro tiveram áreas de cana-de-açúcar colhidas superiores à soma das culturas de soja e de milho.

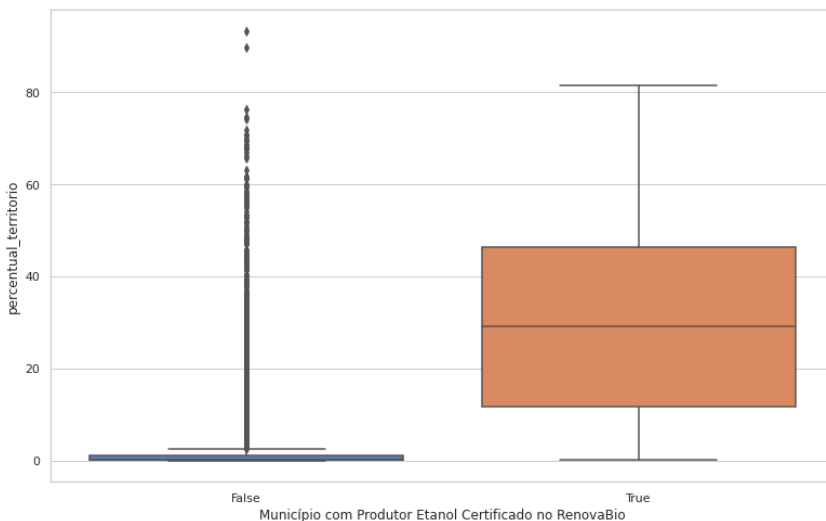
Resultados específicos para a Pergunta 2:

- Há forte indicativo de que a predominância de uma ou de outra cultura em um município constitui importante fator na quantidade de ocorrências de focos de calor.
- As duas mesorregiões analisadas, uma com bioma predominante de Mata Atlântica e outra de Cerrado, indicam que há uma relação entre o percentual da área plantada de cana-de-açúcar em um município com o incremento do índice de focos de calor por unidade de área naquele município, comportamento identificado também na resposta à questão 1.
- A mesma relação entre a presença da cultura no município e o incremento no índice de focos de calor não foi identificada para as outras duas culturas analisadas, milho e soja.

Pergunta 3: Os municípios em que existe produtor de biocombustível certificado no programa RenovaBio têm área de cana-de-açúcar colhida e quantitativo de focos de calor semelhantes ou diferentes dos demais municípios produtores de cana-de-açúcar?

Comparando-se o índice do percentual de área colhida em relação à área total do município entre os grupos de municípios produtores de cana-de-açúcar com e sem usinas de etanol primeira geração certificadas no RenovaBio, verifica-se que a área colhida corresponde a uma parcela consideravelmente maior da área dos municípios quando há usina certificada no programa.

Figura 13 | Índice percentual de Área de Cana Colhida em relação à Área Municipal dos Municípios Produtores de Cana com e sem Usinas Certificadas



O Quadro 6 a seguir detalha numericamente essa diferença entre os dois grupos:

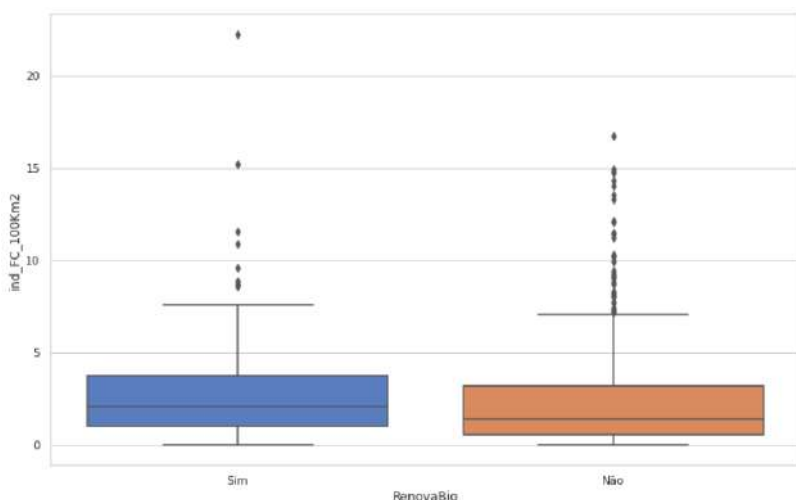
Quadro 6 | Índice Percentual de Área de Cana Colhida em relação à Área Municipal dos Municípios Produtores de Cana com e sem Usinas Certificadas

Município com produtor certificado no RenovaBio?	Área de cana colhida *100 / Área território município		
	Mediana	Média	Desvio padrão
Não	0,10	4,92	12,16
Sim	29,16	31,15	21,22

A Figura 14 apresenta o índice de focos de calor por 100 km² separando os municípios do Brasil que produzem cana-de-açúcar em dois grupos:

municípios com usinas certificadas no RenovaBio e municípios sem a presença de usina certificada, mas com área de cana-de-açúcar colhida em 2020 igual ou superior a 5 km².

Figura 14 | Índice de Focos de Calor em Municípios Produtores com e sem Usina Certificada no RenovaBio



O gráfico apresentado na Figura 14 compreende o total de 967 municípios, sendo que desses, 227 compõem o grupo de municípios com usinas certificadas no programa RenovaBio e 740 fazem parte do grupo de municípios que não possuem usinas certificadas.

A média do índice de focos de calor por 100 Km² nos municípios com usinas certificadas é de 2,74. Já os municípios sem usinas certificadas possuem uma média desse índice de 2,45, conforme dados descritivos do Quadro 7.

Cabe ressaltar que, além de possuir uma quantidade de focos de calor média menor, os municípios sem certificação do RenovaBio também apresentam valores de mediana e de 1º e 3º quartil menores, o que constitui indicativo de que os municípios produtores de cana-de-açúcar sem usinas certificadas apresentam menos focos de calor do que os que possuem usinas certificadas.

Quadro 7 | Resumo Estatísticas Descritivas do Índice de Focos de Calor em Municípios com e sem Usinas Certificadas no RenovaBio

Ind_FC_100 Km ²								
Grupo	Quantidade Municípios	Média	Desvio padrão	Mínimo	1º quartil	Mediana	3º quartil	Máximo
Sim	227	2,74	2,63	0,00	0,98	2,06	3,76	22,21
Não	740	2,45	2,56	0,00	0,54	1,40	3,15	16,72

Resultados específicos para a Pergunta 3:

- Os municípios produtores de cana-de-açúcar em que há usinas certificadas no programa RenovaBio apresentam área de cana-de-açúcar colhida relativa ao seu território expressivamente superior aos demais municípios produtores.
- A média e a mediana do índice de focos de calor por área do município é superior nos municípios que têm usina certificada para etanol primeira geração no RenovaBio do que nos demais municípios produtores de cana-de-açúcar com área colhida maior ou igual a 5 km².

CONCLUSÃO

As análises realizadas demonstraram que os municípios que cultivam cana-de-açúcar têm perfil diferente dos demais quanto à quantidade de focos de calor por km², tendendo esses focos serem mais incidentes quanto maior o índice de área de cana-de-açúcar colhida relativa à área do município. Essa tendência foi encontrada ao se analisar dados, em âmbito nacional, dos biomas Cerrado e Mata Atlântica, em âmbito estadual, analisando-se Goiás e São Paulo e, no nível de mesorregião, considerando as duas selecionadas Norte Central Paranaense e Sul Goiano.

Tendência oposta foi identificada nos municípios quando se analisou sob as perspectivas das áreas de soja e de milho colhidas, as duas culturas temporárias mais presentes no país. Em nível nacional, nos biomas Cerrado e Mata Atlântica, assim como nas duas mesorregiões analisadas, a tendência identificada para essas duas culturas foi de menos focos de calor quanto maior a área relativa colhida.

Em que pese o índice de focos de calor no Cerrado ser mais que o dobro do que o da Mata Atlântica (dados de 2012 a 2021), as tendências de índice de focos de calor por km² em função da área relativa de cana-de-açúcar colhida nos municípios são similares entre os dois grupos de municípios em 2020. O Sul Goiano, mesorregião com amplo predomínio do cerrado, aparece na sexta posição considerando o índice de focos de calor entre as dez maiores mesorregiões produtoras de cana-de-açúcar.

Efetuando-se o recorte dos municípios em que há usinas certificadas em etanol primeira geração no programa RenovaBio conclui-se que, além de eles apresentarem um percentual bem superior de área colhida de cana-de-açúcar comparando-se com os demais municípios produtores de cana, eles registraram índices de focos de calor por km² superiores aos demais municípios produtores com área de cana colhida maior ou igual a 5 km² (valor mínimo dentre os municípios com usinas certificadas).

O RenovaBio e o seu contexto como política pública demanda ainda respostas que podem ser temas de outros estudos. Um dos mais importantes, na perspectiva do desenho da política, é sobre o conceito de adicionalidade, descrito no Artigo 12 do Protocolo de Kyoto, no sentido de que as reduções de emissões certificadas seriam as adicionais às que teriam ocorrido na ausência da atividade de projeto certificada. Qual seria a medida justa de aplicação desse conceito na indústria de biocombustíveis, em especial do etanol, presente no Brasil há décadas? Nesse contexto, haveria espaço para o estabelecimento de metas de redução de emissão para esses produtores? Atualmente, considerando ainda o necessário alinhamento com outras

políticas públicas, seria importante incluir condicionantes e metas mais amplas alinhadas ao espectro ESG¹⁰?

Os resultados obtidos com a pesquisa reforçam a necessidade de se acompanhar a efetividade das políticas públicas de energia e de meio ambiente envolvidas e de avaliá-las de forma integrada, assegurando que o recurso público (ou o aumento de custo imposto ao particular) esteja realmente propiciando as transformações almejadas, efetuando, se necessários, ajustes na execução ou mesmo alterações nos desenhos dessas políticas.

REFERÊNCIAS

BRASIL. Ministério da Ciência, Tecnologia e Inovações. Secretaria de Pesquisa e Formação Científica. Quarta Comunicação Nacional do Brasil à Convenção Quadro das Nações Unidas sobre Mudança do Clima / Secretaria de Pesquisa e Formação Científica. Brasília: Ministério da Ciência, Tecnologia e Inovações, 2021. 620 p.: il. ISBN: 978-65-87432-18-2.

_____. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. **Painel Dinâmico do RenovaBio**. Disponível em: <<https://www.gov.br/anp/pt-br/assuntos/renovabio>>. Acesso em: 24 mar. 2022.

_____. Ministério da Agricultura, Pecuária e Abastecimento. **Dinâmica na cultura da cana-de-açúcar no Brasil: 1990 a 2018**. Jaguariúna: Embrapa Meio Ambiente, PDF 41 p.: il. color. ISSN 1516-4691, 2020.

_____. Perguntas Frequentes. **Queimadas**: Instituto Nacional de Pesquisas Espaciais. Disponível em: <<https://queimadas.dgi.inpe.br/queimadas/portal/informacoes/perguntas-frequentes>>. Acesso em: 10 maio 2022.

10 ESG é uma sigla em inglês que significa *environmental, social and governance* e corresponde às práticas ambientais, sociais e de governança de uma organização. O termo foi cunhado em 2004 em uma publicação do Pacto Global, em parceria com o Banco Mundial, chamada *Who Cares Wins*.

_____. Empresa de Pesquisa Energética. **Estudos do Plano Decenal de Expansão de Energia 2031 - Oferta de Biocombustíveis**. Disponível em: <https://www.gov.br/mme/pt-br/assuntos/secretarias/spe/publicacoes/plano-decenal-de-expansao-de-energia/pde-2031/cadernos/pde-2031-oferta-de-biocombustiveis_06dez2021.pdf/view>. Acesso em: 19 ago. 2022.

WIKIPEDIA. Gases do Efeito Estufa. Disponível em: <https://pt.wikipedia.org/wiki/Gases_do_efeito_estufa>. Acesso em: 10 maio 2022.



**TEMA: SISTEMA ÚNICO
DE SAÚDE**

FERRAMENTA: PROCESSAMENTO
DE LINGUAGEM NATURAL,
DETECÇÃO DE ANOMALIA

ANÁLISE DE SENTIMENTOS E MODELAGEM DE TÓPICOS DE MENSAGENS DO *TWITTER* COMO INSTRUMENTO PARA MONITORAMENTO CONTÍNUO DE POLÍTICAS PÚBLICAS: UM ESTUDO DE CASO DA POLÍTICA PÚBLICA DO SISTEMA ÚNICO DE SAÚDE¹

THIAGO MORAIS FURUCHIMA⁽¹³⁾

INTRODUÇÃO

Redes sociais fazem parte do cotidiano dos brasileiros, seja para compartilhar fotos com amigos e familiares, divulgar ideias, planos e frustrações, manter-se atualizado acerca de determinado assunto, seja para dividir um pouco do cotidiano com o mundo, as redes sociais estão por toda parte.

¹ Este artigo resulta de trabalho de conclusão orientado pelo Professor Alex Lopes Pereira.

Todos os dias, milhares de brasileiros compartilham publicamente informações sobre diversos assuntos, em múltiplos contextos, nas mais variadas plataformas (*Instagram, Facebook, Twitter, TikTok, Youtube e Twitch* estão entre as mais comuns).

Embora as redes sociais sejam extremamente populares no país, pouca atenção tem sido dada à sua utilização no contexto de políticas públicas. Tipicamente, a utilização das redes sociais pela Administração Pública dá-se no sentido de divulgar ou promover ações governamentais, sendo raros os casos de sua utilização para geração sistemática de informação ou obtenção de *insights* para o próprio governo.

Imagine, por exemplo, que uma pessoa ao vacinar-se no posto de saúde de sua cidade compartilhe uma foto no *Instagram* comemorando que está imunizada contra a covid-19; ou um jovem que reclame no *Twitter* que passou o dia inteiro na fila para solicitar o seguro desemprego; ou que a sua mãe esteja há mais de três meses na fila do INSS e ainda não conseguiu agendar a perícia médica; embora fictícios, todos esses exemplos trazem informações que poderiam ser utilizadas para monitorar e avaliar a execução de políticas públicas.

Nesse sentido, o principal objetivo desse projeto foi evidenciar o potencial de utilização de informações públicas disponíveis em redes sociais para o monitoramento de políticas públicas no Brasil.

De forma mais específica, foi desenvolvida uma metodologia para o monitoramento da política pública do Sistema Único de Saúde (SUS) por meio da análise de mensagens compartilhadas na rede social *Twitter*.

Conforme será apresentado, mesmo em um curto período de análise, a metodologia desenvolvida foi capaz de identificar diversas questões relevantes relacionadas à política pública do Sistema Único de Saúde, evidenciadas por meio da comparação dos assuntos identificados, com notícias veiculadas na imprensa.

Ademais, foram observadas questões relevantes que, até a conclusão deste trabalho, não foram noticiadas na imprensa, o que indica que a metodologia desenvolvida é especialmente útil para identificação tempestiva de situações novas, evidenciando sua aplicabilidade no monitoramento contínuo de políticas públicas.

REFERENCIAL TEÓRICO

Tradicionalmente, o desenho de políticas públicas envolve a utilização de dados e estatísticas oficiais coletadas e elaboradas por órgãos e instituições governamentais (SINGH *et al.*, 2020). Contudo, ao longo dos últimos anos, governos e *policy makers* têm percebido a importância da utilização das redes sociais para esse fim (JOSEPH *et al.*, 2017). Conforme apontado por Vydra e Kantorowicz (2021), as redes sociais possuem um potencial enorme de utilização no contexto de políticas públicas.

Nessa seara, conforme ressaltado por Coşkun e Ozturan (2018), diversas pesquisas têm sido realizadas com uso das informações da rede social *Twitter*, tanto em razão da popularidade da plataforma quanto em função da facilidade para realização de coleta de dados da rede, que fornece API específica para essa finalidade.

Dentre os trabalhos que merecem destaque, ressaltamos a pesquisa realizada por Antenucci *et al.* (2014), que utilizou informações da rede social *Twitter* para criar um indicador de empregabilidade nos Estados Unidos. Nesse trabalho, os autores ressaltam que a utilização das redes sociais para criação de indicadores possui as seguintes vantagens: 1) os dados estão disponíveis em tempo real e em uma grande frequência de atualização; 2) as informações obtidas por meio das redes sociais possuem baixo custo, quando comparadas aos tradicionais *surveys*; e 3) as informações oriundas das redes sociais representam o fenômeno natural de comunicação entre indivíduos da sociedade.

Como resultado dessa pesquisa, foi criado o *University of Michigan Social Media Job Loss Index*, que apresentou correlação relevante com os índices oficiais de desemprego do governo norte-americano.

Outra utilização relevante no contexto de políticas públicas é apontada pelos autores Dhimanaarzo e Toshniwaldurga (2021), que destacam a utilização da rede social *Twitter* como alternativa para obtenção de informações tradicionalmente geradas por meio de pesquisas de opinião. Os pesquisadores usaram informações dessa rede social para avaliar a efetividade da campanha *Swachh Bharat Abhiyan* (SBA) de limpeza urbana implementada pelo governo da Índia e mostraram relevante correlação linear ($r = .77$) entre a análise de sentimentos, utilizando as informações da rede social e as pesquisas de opinião realizadas pelo governo indiano.

Merece destaque, ainda, o trabalho desenvolvido por Jafarzadeh *et al.* (2021) que analisou as mensagens compartilhadas no *Twitter* para entender como os cidadãos da Nova Zelândia reagiram à disseminação da covid-19 no país. Neste trabalho, foi utilizada modelagem de tópicos e, dentre as conclusões alcançadas, observou-se que uma das maiores preocupações dos neozelandeses durante o período da pandemia foi com a economia do país.

Paralelamente ao seu potencial uso, Vydra e Kantorowicz (2021) ressaltaram a existência de diversos desafios no uso de informações de redes sociais, precipuamente relacionados ao baixo controle sobre o processo de geração dos dados (respeitadas as regras da própria plataforma de rede social, qualquer pessoa é livre para postar o que quiser), o que demanda cuidados adicionais para seu efetivo uso.

COLETA DE DADOS

Para a obtenção dos dados necessários à realização deste trabalho foi desenvolvido uma aplicação em linguagem de programação *Python*, hospedada em nuvem, a qual fez uso da API (*Application Programming Interface*) disponibilizada pela própria plataforma *Twitter*.

A referida aplicação realizou a coleta contínua (24/7) de mensagens que mencionaram as palavras-chave “SUS” ou “Sistema Único de Saúde”. Todas as mensagens publicamente compartilhadas por usuários da plataforma que continham quaisquer desses termos foram filtradas e armazenadas em banco de dados para realização desta pesquisa.

Seguindo essa abordagem, entre maio e agosto de 2022, foram coletadas cerca de 230 mil mensagens distintas relacionadas ao tema de interesse.

Como será apresentado na próxima seção, essas mensagens foram utilizadas como *proxy* para obtenção do sentimento da população brasileira em relação à política pública do Sistema Único de Saúde, bem como para a identificação de questões importantes no contexto do monitoramento contínuo dessa política.

Importante destacar que existe um viés de seleção inerente a essa abordagem: grande parte da população brasileira não faz uso do *Twitter* (muitos sequer utilizam qualquer rede social). Além disso, se considerarmos a idade dos usuários de redes sociais no Brasil, podemos constatar que parte considerável da população brasileira não estará adequadamente representada na amostra selecionada.

Entendemos que, embora relevantes, tais limitações não são impedimento à utilização das informações para o que se propõe neste projeto, especialmente, se considerarmos a popularização crescente das redes sociais, entre os mais diversos públicos.

ANÁLISE DE SENTIMENTOS

A análise de sentimentos de mensagens da rede social *Twitter* já foi objeto de pesquisa de diversos atores, sendo considerada uma área especializada no campo de linguística computacional. Nesse contexto, a análise de sentimentos busca identificar e avaliar as opiniões expressadas pelos usuários da rede social, utilizando métodos automatizados (ZIMBRA *et al.*, 2018).

Conforme apontado pelo mesmo autor, diversos métodos de aprendizagem de máquina podem ser empregados para a realização dessa tarefa, com destaque para os algoritmos de Regressão Logística (*Logistic Regression*), Máquinas de Vetores de Suporte (*Support vector machines - SVM*) e Redes Neurais Artificiais (*Artificial Neural Network*), as quais representam hoje o estado da arte no que tange à análise de sentimentos, sendo essa a abordagem utilizada neste trabalho.

Para isso, foi utilizado o modelo *XLM-T* apresentado em (BARBIERI, ESPINOSA ANKE; CAMACHO-COLLADOS, [s.d.]). Trata-se de um modelo do tipo *Bidirectional Encoder Representations from Transformers (BERT)* que foi treinado em aproximadamente 198 milhões de *tweets* e ajustado especificamente para a realização da tarefa de análise de sentimentos. Esse modelo foi disponibilizado pelos autores sob a licença *Open Source Apache License 2.0*, tendo sido utilizado sem ônus neste trabalho.

A tabela abaixo apresenta amostra de mensagens reais, após a análise de sentimentos, utilizando o supracitado modelo *XLM-T*. As informações que possivelmente poderiam ser utilizadas para identificação do autor da mensagem foram suprimidas.

Tabela 1 | Amostra de mensagens e sentimento calculado pelo modelo XLM-T

Texto	Sentimento	Score
Viva o SUS! Ops, Viva o Sírio Libanês!	<i>Positive</i>	0.579726
#RolTaxativoMata NOTA PÚBLICA: rol taxativo é um retrocesso contrário à vida e ao SUS.	<i>Negative</i>	0.805813
Que matéria linda e comovente no Fantástico. A emoção do médico, as cenas dos irmãos se tocando... Tem emoções que só mesmo a ciência e o SUS podem nos proporcionar.	<i>Positive</i>	0.906965
O SUS tem que acabar. É questão de humanidade. Privatiza-lo é a melhor alternativa.	<i>Negative</i>	0.826747

Texto	Sentimento	Score
assisti um filme completo esperando a cirurgia simplesmente amo o SUS.	<i>Positive</i>	0.823064
Mais um mês pagando 560 reais nos meus remédios, remédios esse que o SUS deveria ter, mas eu NUNCA encontrei.	<i>Negative</i>	0.556566
Mostre uma monstruosidade maior do que roubar bilhões que deveriam ser usados na educação e na saúde, onde, no governo petista, o SUS era um descalabro! Não precisa muito, busca no Google, e veja os corredores dos hospitais na era PT. Aproveite e veja a classificação no Pisa!	<i>Negative</i>	0.783703
“O documento prevê que a classe média deve pagar mensalidades nas universidades públicas e pelo atendimento no SUS” - e a #classemedia apoiando cegamente um grupo da elite nacional q entende que a nação serve para atendê-la...	<i>Neutral</i>	0.687669
4° dose veio aí, amém SUS 🙌🙌🙌🙌	<i>Positive</i>	0.815490
Perdão. Viva o SUS, mas tô rindo feito um trenó descarrilhado.	<i>Negative</i>	0.824573
Agora pronto venderam o SUS pro elon musk.	<i>Neutral</i>	0.634898

Em que pese tratar-se de um modelo *state-of-the-art*, é possível observar pelos exemplos acima que a análise automatizada não é perfeita. Textos em que o autor se vale da ironia, como o observado em “*Viva o SUS! Ops, Viva o Sírio Libanês!*” por exemplo, representam ainda um desafio para a análise de sentimentos. Embora a análise automatizada tenha classificado esse texto como positivo, é evidente que o autor, na verdade, queria expressar exatamente o contrário.

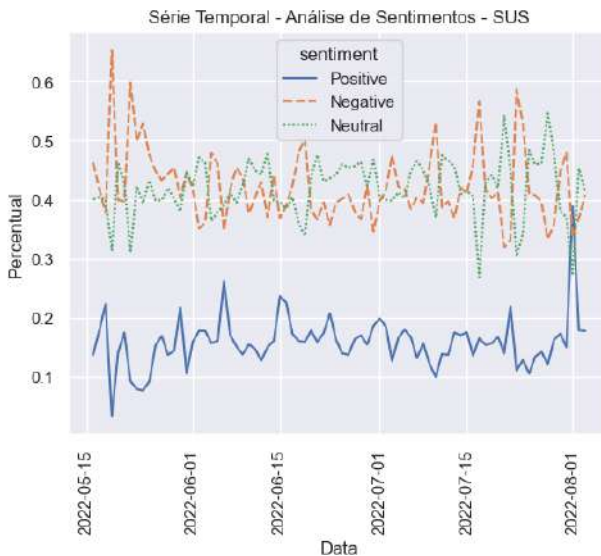
Outro exemplo em que a classificação não foi adequadamente realizada pode ser observada em “*Perdão. Viva o SUS, mas tô rindo feito um trenó descarrilhado*”. Em que pese esse texto ter um caráter positivo, a análise automatizada classificou o texto como negativo.

A despeito dessas questões, a análise realizada em amostra aleatória mostra que, de maneira geral, o modelo consegue extrair adequadamente o sentimento das postagens analisadas.

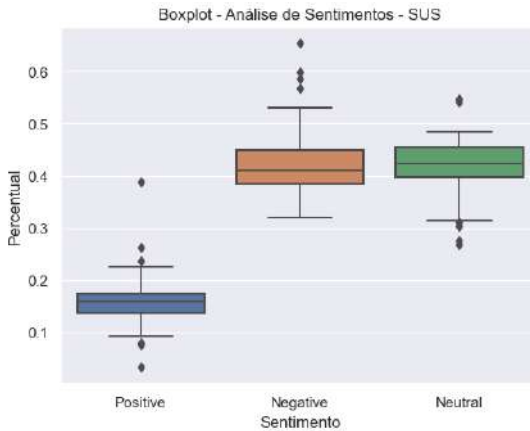
Além disso, tendo em vista que neste trabalho o objetivo foi identificar o sentimento geral em determinado período, divergências pontuais não possuem o potencial de impactar negativamente o resultado geral, especialmente se consideramos a população analisada: 230 mil mensagens. Destaca-se, contudo, que essas divergências pontuais não devem ser desconsideradas em amostras pouco representativas.

O gráfico abaixo representa a variação temporal diária do sentimento geral calculado a partir das mensagens analisadas que mencionaram as palavras “SUS” e “Sistema Único de Saúde”.

Figura 1 | Série Temporal - Análise de Sentimentos - SUS



Como pode ser observado, em geral as mensagens compartilhadas na rede social *Twitter* que mencionaram os termos de interesse possuem sentimento geral neutro (mediana 0.42), seguidos pelo sentimento negativo (mediana .41) e positivo (mediana .15). O gráfico a seguir mostra de forma mais clara essa variação.

Figura 2 | Boxplot - Análise de Sentimentos - SUS

A partir desses gráficos, fica evidente também que há grande variação do sentimento (positivo, negativo e neutro) ao longo do tempo. O sentimento negativo, por exemplo, possui valor máximo de 0.65. Isso significa que, em determinado dia, aproximadamente 65% das mensagens compartilhadas na rede social sobre o SUS possuíam caráter negativo. Similarmente, existem casos em que o sentimento positivo foi o prevalente em aproximadamente 38% das mensagens da rede social sobre o assunto no dia.

Esses picos no gráfico diário representam dias atípicos e são de especial relevância para este trabalho, pois podem trazer informações relevantes para o monitoramento da política pública analisada.

Para a identificação automática dessas datas atípicas (em contrapartida à identificação visual por meio de inspeção gráfica), foi utilizado método para identificação de anomalias (*anomaly detection*). De forma mais específica, foi utilizado algoritmo de aprendizado de máquina não supervisionado *Isolation Forest* (TONY LIU; MING TING; ZHOU, [s.d.]), disponível na biblioteca *Python Scikit-Learn* (PEDREGOSA *et al.*, 2011) recebendo como entrada as informações dos sentimentos gerais computadas em cada dia e gerando como resultado as datas consideradas *outliers*.

A tabela a seguir apresenta o resultado da aplicação desse algoritmo para identificação dos dias atípicos:

Tabela 2 | Datas identificadas como atípicas pelo algoritmo de detecção de anomalias

Data	Positive	Negative	Neutral
2022-05-19	0.032161	0.653516	0.314322
2022-05-22	0.091678	0.598025	0.310296
2022-05-23	0.078497	0.498573	0.422931
2022-05-24	0.076164	0.529972	0.393865
2022-06-06	0.261413	0.347545	0.391042
2022-07-17	0.164948	0.567010	0.268041
2022-07-23	0.109890	0.586813	0.303297
2022-08-01	0.388332	0.337284	0.274385

Como pode ser observado, comparando a tabela acima com o gráfico da série temporal dos sentimentos diários, o algoritmo consistentemente aponta como atípicos os picos positivos e negativos da série temporal.

A identificação dessa relevante variação evidencia que existe uma mudança sistemática do sentimento geral das mensagens compartilhadas na rede ao longo do tempo. Não se tratam de mensagens casualmente compartilhadas por usuários específicos, mas de centenas (às vezes milhares) de pessoas que, por alguma razão, publicaram na rede social opiniões coincidentes. Tal situação nos leva a seguinte pergunta: a partir dessas diversas mensagens compartilhadas, seria possível apontar o que motivou esses picos no sentimento geral das mensagens sobre determinado assunto?

MODELAGEM DE TÓPICOS

Com o objetivo de tentar identificar os assuntos possivelmente relacionados aos picos positivos e negativos nos sentimentos computados, conforme

apresentado anteriormente, foi realizada a modelagem de tópicos utilizando a técnica *Latent Dirichlet Allocation* (LDA), proposta por (BLEI; NG; EDU, 2003). Conforme apresentado no trabalho deles, a LDA tenta descobrir a estrutura subjacente de uma coleção de documentos tratando cada documento como sendo gerado por uma mistura de diferentes tópicos. No caso específico deste trabalho, essa técnica foi empregada para tentar identificar assuntos comuns nas mensagens compartilhadas nas datas em que o sentimento geral computado foi identificado como atípico, conforme metodologia apresentada anteriormente.

Para essa tarefa, foi utilizada a biblioteca *Python Gensim* (ŘEH UŘEK; SOJKA, 2010), também disponibilizada sob licença de código aberto. Conforme documentação oficial, a *Gensim* é uma biblioteca criada especificamente para as tarefas de modelagem de tópicos, a indexação de documentos e a descoberta de similaridade em grandes bases textuais.

Dessa forma, para os dias em que o sentimento geral computado pela análise de sentimentos foi atípico, realizou-se a modelagem de tópicos das mensagens daquela data. Os resultados e as discussões serão apresentados na próxima seção.

PRINCIPAIS RESULTADOS

A partir das informações dos sentimentos diários computados pela análise de sentimentos, e dos dias atípicos identificados pelo método de identificação de anomalias, foi realizada a modelagem de tópicos das mensagens.

19 de maio de 2022

Para o dia 19 de maio de 2022, a modelagem de tópicos identificou como mais relevante o tópico com as seguintes palavras: [“bolsolão”, “brasil”, “gasta”, “paga”, “corrupção”, “dia”, “governo”, “escândalo”, “agora”, “mec”].

Em pesquisa realizada no *Google* pelos termos “Bolsolão” e “SUS”, incluindo como parâmetro a data 19/05/2022, foram identificadas diversas notícias

veiculadas na imprensa sobre um possível desvio de recursos do Fundo Nacional de Saúde para favorecimento de aliados do presidente Bolsonaro. Abaixo é apresentada manchete referente ao tema:

Figura 3 | Manchete de notícia sobre o assunto —19/5/2022

Dinheiro destinado ao SUS é usado para beneficiar aliados do governo no Congresso

Fundo Nacional de Saúde (FNS) distribuiu em 2021 boa parte dos R\$ 7,4 bilhões em emendas de relator a redutos eleitorais de caciques do Centrão

Patrik Camporez, Eduardo Gonçalves, Aguirre Talento, Paula Ferreira e Dimitrius Dantas
15/05/2022 - 04:30 / Atualizado em 15/05/2022 - 09:06

22, 23 e 24 de maio de 2022

Para os dias 22, 23 e 24 de maio de 2022, a modelagem de tópicos identificou como mais relevante o tópico com as seguintes palavras: [“militares”, “fim”, “projeto”, “gratuidade”, “poder”, “prevê”, “manter”, “querem”, “cobrança”, “pública”].

Em pesquisa realizada no *Google* pelos termos “Militares” e “SUS”, na data 24/5/2022, foram identificadas diversas matérias com a seguinte manchete: “Projeto de militares prevê fim do SUS gratuito em 2035”.

Figura 4 | Manchete de notícia sobre o assunto — 22, 23 e 24/5/2022

 Estadão

Projeto de militares prevê manter poder até 2035 e fim da gratuidade no SUS em 2025

O documento prevê que a classe média deve pagar mensalidades nas universidades públicas e pelo atendimento no SUS. A cobrança deve começar em 2025.

May 23, 2022



6 junho de 2022

Para o dia 6 de junho de 2022, a modelagem de tópicos identificou como mais relevante o tópico com as seguintes palavras: ["viva", "susguga", "sírio", "libanês", "mim", "lula", "sírio", "hospital", "porque", "povo"].

Não foi possível identificar notícias veiculadas na imprensa com os assuntos supracitados. Contudo, uma análise detalhada das mensagens indica que o pico negativo nesse dia pode estar relacionado à manifestação pública de indignação sobre o fato de que o ex-presidente Luiz Inácio Lula da Silva, ao ser diagnosticado com covid-19, utilizou o hospital particular Sírio Libanês, em vez da rede pública de saúde. A mensagem abaixo ratifica essa suposição:

Figura 5 | Tweet publicado sobre o assunto — 6/6/2022



17 de julho de 2022

Para o dia 17 de julho de 2022, a modelagem de tópicos identificou como mais relevante o tópico com as seguintes palavras: ["tira", "dinheiro", "mec", "propina", "arthur", "paga", "lira", "deputados", "centrão", "assim"].

Em pesquisa realizada no *Google* pelos termos anteriores, não foi possível identificar notícia veiculada na imprensa, na referida data, sobre possível esquema de corrupção no SUS envolvendo deputados do centrão e Arthur Lira.

Destaca-se, contudo, que esse pode ser exatamente um achado deste trabalho. O fato de a análise de sentimentos e a modelagem de tópicos terem identificado como relevantes os termos supracitados representam subsídios importantes para um aprofundamento.

23 de julho de 2020

Para o dia 23 de julho de 2022, a modelagem de tópicos identificou como mais relevante o tópico com as seguintes palavras: [“políticos”, “rj”, “cirurgias”, “lotearam”, “troca”, “sisreg”, “votos”, “denuncia”, “viva”, “ciência”].

Em que pese as palavras-chave identificadas pela modelagem de tópicos indicarem uma possível denúncia envolvendo “políticos do RJ”; “cirurgias”; “troca de votos”; e o “Sisreg” (Sistema Nacional de Regulação), não foram identificadas notícias que tratam do assunto.

Assim, como ocorrido no dia 17/6/2022, esse pode ser um achado deste trabalho, indicando questão relevante, relacionada a possíveis condutas irregulares na gestão do SUS no Rio de Janeiro.

Embora esses tópicos não possam ser tratados por si só como evidências, podem ser utilizados como subsídios para uma investigação ou, no mínimo, um aprofundamento dentro de um processo de monitoramento contínuo da política pública.

1º de agosto de 2022

Por fim, para o dia 1º de agosto de 2022, único dia em que o sentimento geral positivo foi superior ao negativo, a modelagem de tópicos identificou como mais relevante o tópico com as seguintes palavras: [“viva”, “gêmeos”, “gente”, “ciência”, “coisa”, “reportagem”, “linda”, “separação”, “bernardo”, “médico”].

A busca por notícias envolvendo as palavras do tópico, para o dia 1º/8/2022, identificou reportagem divulgada no telejornal “Fantástico”, sobre um importante marco para a ciência, e, em especial, para a medicina brasileira.

Figura 6 | Manchete de notícia sobre o assunto — 1º/8/2022

Gêmeos unidos pelo crânio são separados após nove cirurgias, e equipe médica brasileira se torna referência

Tratamento médico foi custeado pelo SUS e o médico inglês Owase Jeelani, referência neste tipo de procedimento, participou de procedimentos. Irmãos dividiram 15% de cérebro e veia importante que conduzia o sangue de retorno aos corações.

Por Fantástico

31/07/2022 22h54 · Atualizado há um mês



CONCLUSÃO

A principal contribuição deste trabalho foi evidenciar o potencial de utilização de informações públicas disponíveis em redes sociais para o monitoramento de políticas públicas no Brasil.

Nesse contexto, a política pública do Sistema Único de Saúde foi utilizada como caso concreto para o estudo e a aplicação de técnicas de Processamento de Linguagem Natural, notadamente a Análise de Sentimentos e a Modelagem de Tópicos, para a obtenção de *insights* importantes para o monitoramento contínuo da política.

A análise de sentimentos permitiu identificar relevante variação no sentimento geral dos brasileiros em relação ao SUS, bem como observar que acontecimentos importantes relacionados à política afetam diretamente a opinião da população. A modelagem de tópicos permitiu apontar possíveis causas para os picos na variação do sentimento geral ao longo do tempo, identificando tanto questões de conhecimento comum, noticiadas na imprensa brasileira, bem como temas novos, potencialmente importantes no contexto geral da política pública, os quais podem ser objeto de aprofundamento pelas instâncias competentes.

Nesse sentido, entendemos que a abordagem utilizada neste trabalho pode ser efetivamente útil para o monitoramento de políticas públicas no Brasil. A tempestividade em que as informações são compartilhadas nas redes sociais e a abrangência do seu alcance tornam a metodologia desenvolvida especialmente atrativa no monitoramento contínuo de políticas públicas.

Ressaltamos também o baixo custo para implementação e manutenção de uma solução automatizada como a aqui apresentada. Enquanto uma pesquisa de opinião com cerca de mil pessoas custa em média entre R\$ 5 mil e R\$ 20 mil, o monitoramento contínuo conforme apresentado neste trabalho (que analisou mais de 200 mil mensagens compartilhadas por mais de 50 mil pessoas) custou cerca de R\$ 150 por mês, resultando em um custo total de aproximadamente R\$ 600².

Por fim, destacamos que a metodologia apresentada aqui (que consiste resumidamente na análise de sentimentos, na detecção das datas em que o sentimento computado foi considerado anômalo e na modelagem de tópicos das mensagens compartilhadas nessas datas) possui possibilidades de melhoria que merecem evidência:

O modelo *deep learning* utilizado para análise de sentimentos, em que pese se tratar de um modelo “estado da arte” nessa tarefa, pode ser melhorado por meio de treinamento específico para a análise de sentimentos do contexto do problema, no caso, monitoramento de políticas públicas, o que aumentaria a acurácia das classificações de forma geral.

A modelagem de tópicos com a utilização da técnica LDA poderia ser melhorada com a utilização de métrica objetiva para a seleção da quantidade de tópicos. Além disso, o emprego de n-gramas para identificação de grupos de palavras nos tópicos (em vez de apenas palavras individuais como feito neste trabalho) tem o potencial de identificar assuntos com mais precisão.

A despeito desses pontos de melhoria, entendemos que os objetivos deste projeto foram atingidos e que o uso das redes sociais para o monitoramento

2 Servidor virtual em nuvem, com CPU compartilhada, 4GB de memória RAM e 100 GB de espaço em disco.

contínuo, e quem sabe até mesmo o desenho de políticas públicas, esteja apenas começando.

REFERÊNCIAS

ANTENUCCI, D. *et al.* **Using Social Media to Measure Labor Market Flows.** mar. 2014.

BARBIERI, F.; ESPINOSA ANKE, L.; CAMACHO-COLLADOS, J. **XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond.** [s.d.].

BLEI, D. M.; NG, A. Y.; EDU, J. B. Latent Dirichlet Allocation Michael I. Jordan. **Journal of Machine Learning Research**, v. 3, p. 993–1022, 2003.

COŞKUN, M.; OZTURAN, M. #europehappinessmap: A framework for multi-lingual sentiment analysis via social media big data (a Twitter case study). **Information (Switzerland)**, v. 9, n. 5, 24 abr. 2018.

DHIMANAARZOO; TOSHNIWALDURGA. A Twitter Framework to assess the Effectiveness of Indian Government Campaign. **Transactions on Asian and Low-Resource Language Information Processing**, 3 abr. 2021.

JAFARZADEH, H. *et al.* Making sense of COVID-19 over time in New Zealand: Assessing the public conversation using Twitter. **PLOS ONE**, v. 16, n. 12, p. e0259882, 1 dez. 2021.

PEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

ŘEH UŘEK, R.; SOJKA, P. **Software Framework for Topic Modelling with Large Corpora.** Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. Anais...Valletta, Malta: ELRA, maio 2010.

SINGH, P. *et al.* Smart Monitoring and Controlling of Government Policies Using Social Media and Cloud Computing. **Information Systems Frontiers**, v. 22, n. 2, p. 315–337, 1 abr. 2020.

TONY LIU, F.; MING TING, K.; ZHOU, Z.-H. **Isolation Forest**. [s.d.].

VYDRA, S.; KANTOROWICZ, J. Tracing Policy-relevant Information in Social Media: The Case of Twitter before and during the COVID-19 Crisis. **Statistics, Politics and Policy**, v. 12, n. 1, p. 87–127, 1 jun. 2021.

ZIMBRA, D. *et al.* The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation. **ACM Transactions on Management Information Systems**, v. 9, n. 2, 1 abr. 2018.



**TEMA: CADASTRO ÚNICO
PARA PROGRAMAS SOCIAIS**

FERRAMENTA:
ESTATÍSTICA DESCRITIVA

ANÁLISE DO CADASTRO ÚNICO (CADÚNICO): UM RECORTE DE ANTES E DEPOIS DA PANDEMIA DA COVID-19¹

CIRO JONATAS DE SOUZA OLIVEIRA, GLAUBER VOLKMER, JOSÉ LEONARDO RIBEIRO NASCIMENTO, LEONARDO LINS CAMARA MARINHO, MARCOS CÂNDIDO DE PAULA REZENDE, MONICA RONDINA, ROMUALDO ANSELMO DOS SANTOS E SÉRGIO LUIZ GOMES DE CARVALHO

INTRODUÇÃO

É de conhecimento comum que o impacto da pandemia do novo Coronavírus foi gigantesco em todo o mundo. Dados da Johns Hopkins University dão conta que até setembro de 2022, mais de 618 milhões de casos foram contabilizados, com registros de óbitos superiores a 6,5 milhões (CORONAVIRUS RESOURCE CENTER, 2022). Seja do ponto de vista da saúde, da economia, seja do aspecto emocional, não há quem tenha passado incólume pelos efeitos da covid-19. As populações mais vulneráveis, contudo, foram as mais atingidas, especialmente no contexto econômico. No Brasil, é possível avaliar se a situação efetivamente aconteceu dessa maneira a partir da análise dos dados do Cadastro Único (CadÚnico), que é um registro que consolida as famílias de baixa renda no Brasil.

1 Este foi um trabalho realizado durante o curso, no Laboratório de Casos de Políticas Públicas Sociais, sob a orientação dos professores: Jackson De Toni, Natalia Massaco Koga, Helio Bomfim de Macedo Filho, Jeniffer de Paula Chaves.

Diante disso, o presente estudo analisa a base de dados do Cadastro Único a partir do perfil das famílias residentes em municípios brasileiros com população superior a 500 mil habitantes. O objetivo geral da pesquisa foi identificar eventuais mudanças no perfil da pobreza e da extrema pobreza nesses grandes municípios no período compreendido entre os anos de 2020 e 2021. A pesquisa, além do foco nos grupos familiares, considerou dados unicamente dos responsáveis dessas famílias, desconsiderando detalhes dos demais componentes dos grupos familiares. O recorte justifica-se em razão do grande volume de dados do CadÚnico e em nome da precisão do estudo.

Como objetivos específicos, o estudo buscou verificar se o CadÚnico tem a capacidade de acompanhar as mudanças sociais e se ele traduz-se em instrumento de referência para as políticas públicas, sobretudo as políticas de transferência de renda. Assim, o estudo assenta-se na seguinte questão de pesquisa: o CadÚnico é instrumento suficiente para revelar a mudança de perfil das famílias que vivem na pobreza e na extrema pobreza nos grandes municípios brasileiros durante a pandemia da covid-19?

REFERENCIAL TEÓRICO

Nesta seção serão desenvolvidos os principais conceitos teóricos a respeito da pobreza e das políticas de transferência de renda, considerando as particularidades do presente estudo. Ademais, serão feitas abordagens sobre a pandemia e o CadÚnico.

POBREZA

A pobreza é algo que se caracteriza muito mais pela experiência do que pelo conceito. Nesse contexto, há múltiplas visões sobre o que seria pobreza e dentro dela existem diferentes extratos que flutuam geograficamente. Rocha (2011), em uma tentativa de resumir a questão, assim descreveu o problema:

Pobreza é um fenômeno complexo, podendo ser definido de forma genérica como situação na qual as necessidades não são atendidas de forma adequada. Para operacionalizar essa noção ampla e vaga, é essencial especificar que necessidades são essas e qual nível de atendimento pode ser considerado adequado. A definição relevante depende basicamente do padrão de vida e da forma como as diferentes necessidades são atendidas em determinado contexto socioeconômico. Em última instância, ser pobre significa não dispor dos meios para operar adequadamente no grupo social em que vive (ROCHA, 2011, p. 9-10).

Dessa maneira, a pobreza está associada ao atendimento de necessidades básicas das pessoas, as quais vão variar conforme as demandas pessoais e os grupos sociais que elas pertencam, ou ainda, as regiões geográficas, como é o caso local em que o problema, ainda que distribuído em todo o território brasileiro, é mais presente nos estados das regiões Norte e Nordeste (SOUZA; COUTO, 2015).

Pode-se depreender ainda do excerto anterior da lavra de Rocha (2011), que a pobreza possui diferentes origens, as quais são de fundamental identificação para tratar o problema. De acordo com as lições de Tomazini (2017), a pobreza tem origens distintas, podendo estar em falha na divisão da riqueza nacional, na insegurança alimentar ou na incapacidade de geração de renda. Segundo o mesmo autor, os programas de transferência de renda brasileiros foram criados com base nessa última origem.

Indicadores de pobreza, tais como baixa escolaridade e localização geográfica, potencializados pela falta de acesso à educação básica e ao técnico-profissional, refletem na característica dos postos de trabalho que a população pobre consegue acessar, os quais são geralmente precários, informais, rotativos e, majoritariamente, degradantes (PRÓSPERI, 2015).

De acordo com o estudo “Mapa da Nova Pobreza”, elaborado pela Fundação Getúlio Vargas Social (FGV Social), o contingente de pessoas com renda domiciliar per capita de até R\$ 497 mensais (ou seja, na linha da pobreza – U\$ 5,5/dia) no Brasil, alcançou mais de 62 milhões de indivíduos no ano de 2021 (NERI, 2022). Outros aspectos para caracterizar a pobreza e aliados à renda podem ser destacados com base nas condições de vida das pessoas, a saber: habitação (acesso à água canalizada, à rede de esgotos e à energia elétrica) e consumo (acesso a fogão, à televisão e à geladeira) (CALDAS; SAMPAIO, 2015, p. 77). Esses autores classificaram domicílios pobres em 5 (cinco) categorias, englobando renda e demais dimensões de consumo e habitação, chegaram à conclusão de que o uso exclusivo da renda como forma de mensurar a pobreza não é mais adequado.

Assim, Caldas e Sampaio (2015, p. 79) entendem que pelo contraste entre renda e outras dimensões pode-se considerar dois grandes tipos de erros nas mensurações de pobreza e nas políticas sociais: o de ter renda, mas apresentar carências nas outras dimensões, e o de não ter renda suficiente, mas não apresentar carências em termos de consumo e serviços de habitação.

Em se tratando de políticas de transferência de renda para combate à pobreza, essas precisam estar atentas às características individuais, tais como idade, sexo, condição de maternidade etc. Rossi (2017), em pesquisa sobre o assunto, identificou que há comportamento distintos entre grupos formados por jovens dos sexos masculino e feminino, jovens mães e não mães, sendo que há maior vulnerabilidade das mulheres, estando essas fora do trabalho em maior número, devido à falta de vagas em creches e pré-escolas.

Nesse sentido, delimitar um perfil de pobreza é essencial para descrever características da população pobre e suas privações em outras dimensões além da renda (OSÓRIO *et al.*, 2011). Isso porque, com esses parâmetros, é possível demonstrar quais os extratos em que há maior vulnerabilidade e, portanto, requerem maior atenção do poder público e formuladores de políticas públicas. Um exemplo da importância desse detalhamento são os estudos que indicam que domicílios chefiados por mulheres possuem maior probabilidade de serem pobres (MARINHO; LINHARES; CAMPELO, 2011).

PANDEMIA DA COVID-19

A Pandemia da covid-19 foi reconhecida pela Organização Mundial de Saúde (OMS) como tal em março de 2020. Desde então, várias medidas foram adotadas pelos diferentes países para conter o avanço da doença e para proteger a população, uma vez que a doença é de transmissão e contaminação rápidas (MALTA *et al.*, 2020). No Brasil, essas medidas variaram de estado para estado e de município para município. De acordo com Malta *et al.* (2020), essas medidas foram o fechamento de escolas e comércios não essenciais, restrição de circulação de pessoas, horário reduzido e(ou) fechamento de comércios, chegando até ao bloqueio total (*lockdown*) em certos casos.

As medidas necessárias de restrição social trouxeram consequências para certas parcelas da sociedade que dependiam justamente da circulação de pessoas, em particular os ambulantes, aqueles com empregos informais e temporários (MATTA *et al.*, 2021). Sendo assim, é possível considerar que a renda dessas pessoas foi severamente comprometida, com reflexos na condição de pobreza e demandando respostas do poder público e dos programas sociais, em particular.

CADÚNICO

O Cadastro Único para Programas Sociais, ou simplesmente CadÚnico, é uma ação federal, com gestão compartilhada e descentralizada entre a União, os estados, o Distrito Federal e os municípios, que contém informações sobre as famílias brasileiras de baixa renda e permite ao governo conhecer melhor a realidade socioeconômica dessa população (MINISTÉRIO DA CIDADANIA, 2022). Em tese, o CadÚnico deveria ser a base de referência para os programas sociais, uma vez que capta um amplo conjunto de informações pessoais (documentação, escolaridade, trabalho e renda) e domiciliares (composição familiar, características do domicílio, acesso a serviços de utilidade pública), fundamentais para a implementação e gestão de políticas sociais nos três níveis da federação” (PAIVA *et al.*, 2020).

O CadÚnico coleta e disponibiliza informações atualizadas periodicamente para 20 programas sociais federais a respeito de um amplo quadro de informações socioeconômicas de 27 milhões de famílias, quase 80 milhões de pessoas de baixa renda, isso é, famílias com rendimento menor ou igual a meio salário-mínimo. (DIREITO; KOGA, 2020, p. 1287).

De acordo com Paiva *et al.* (2020), as extrações do Cadastro Único, realizadas mensalmente pela Caixa Econômica Federal (Caixa) e pelo Ministério da Cidadania, são o mais rico registro administrativo para estudos sobre políticas sociais. Ainda que seja uma importante iniciativa e ferramenta, o CadÚnico não está ausente de críticas. Segundo Direito e Koga (2020, p. 1303), a centralidade do CadÚnico como receptor de dados de programas sociais “é praticamente nula” o que sugere que “não há uma retroalimentação ou integração dos dados que permita ajustes mais precisos sobre a situação das famílias, o que poderia aprimorar a implementação de todas as políticas”. As pesquisadoras concluem que existe distância significativa entre os beneficiários e a gestão da maior parte dos programas e da gestão central do cadastro, o que também inviabilizaria uma melhor implementação dessas políticas.

Em linha com essa posição crítica, Botelho *et al.* (2020) argumentam que um programa de proteção social eficaz precisa ter agilidade para reconhecer rapidamente a mutação da condição socioeconômica de uma família. Esse reconhecimento, por óbvio, passa por registros ágeis e adequados, como potencialmente o CadÚnico pode transformar-se. Em suma, o que Botelho *et al.* (2020, p. 6-7) advogam é que quem cair na pobreza e quem sair dela, devem, respectivamente, ser habilitados e desabilitados rapidamente, e que quando a família for emancipada da condição de pobreza, ela deve dar lugar a outra família mais necessitada de maneira célere.

Essa ideia já está contida no CadÚnico, a dificuldade é torná-lo célere o suficiente para identificar essas mudanças de perfis e acompanhar as flutuações sociais de maneira tempestiva. O aperfeiçoamento e a intensificação do uso do Cadastro Único para Programas Sociais (CadÚnico) representam parte es-

sencial da construção de uma política de proteção social eficaz. Um CadÚnico mais abrangente e ágil será capaz de fazer a devida identificação das famílias e suas carências principais (BOTELHO *et al.*, 2020, p. 7).

METODOLOGIA

Para o presente estudo, foram feitos os seguintes recortes na base de dados do CadÚnico:

- Apenas municípios com mais de 500 mil habitantes.
- Na análise dos dados individuais, foram considerados apenas os dados dos responsáveis familiares dos municípios com mais de 500 mil habitantes.

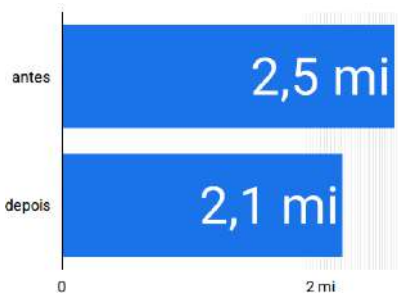
Os cadastros foram divididos em dois grupos:

Antes da pandemia — Pessoas cadastradas antes de 14/3/2020 e com cadastro atualizado depois dessa data.

Depois da pandemia — Pessoas cadastradas depois de 14/3/2020, data de início da pandemia.

A figura 1 a seguir mostra a quantidade de cadastros compreendidas no recorte do presente trabalho:

Figura 1 | Quantidade de pessoas cadastradas antes e depois da pandemia e que tiveram dados atualizados nesse período em municípios com mais de 500 mil habitantes



Fonte: Elaborado pelo autor.

O mapa da Figura 2 mostra a concentração das cidades com população acima de 500 mil habitantes:

Figura 2 | Distribuição dos municípios com mais de 500 mil habitantes e concentração dos cadastrados considerados no estudo



Fonte: Elaborado pelos autores.

A partir dos recortes indicados e da divisão dos cadastros nos grupos antes da pandemia e depois da pandemia, foram realizados vários testes comparativos, tais como:

- Idade média das pessoas cadastradas.
- Proporção de pessoas do sexo masculino e feminino.
- Proporção de alfabetizados.
- Proporção de pessoas que frequentam ou frequentaram escola.
- Proporção de pessoas por grau de instrução.
- Proporção de pessoas por aspectos ocupacionais.

- Proporção de pessoas por faixa de renda.
- Proporção de pessoas no Cadastro Único que recebem o Auxílio Brasil.
- Tamanho da família em que o responsável familiar é um homem.

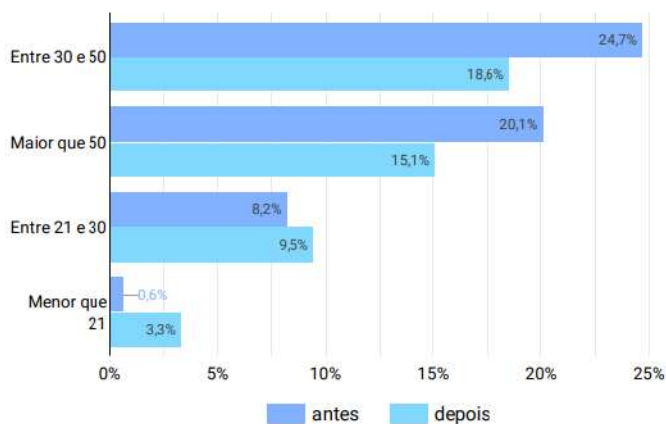
RESULTADOS

Após as análises dos dados, verificou-se que a idade média das pessoas inscritas no CadÚnico caiu após a pandemia e que a proporção de jovens abaixo de 21 anos aumentou cinco vezes em relação ao período anterior à pandemia. Houve grande migração de pessoas do sexo masculino para o CadÚnico após a pandemia, visto que a proporção de homens inscritos como responsável familiar quase dobrou, passando de 14% para 27%. A proporção de pessoas que compõem o CadÚnico e que possuem nível superior, dentro do recorte especificado, dobrou após a pandemia.

A partir dos resultados verificados, percebe-se a necessidade de adotar algumas medidas, tais como: a criação de um programa específico para requalificação e atualização de pessoas com nível superior, partindo da premissa de que, num primeiro momento de retomada do crescimento econômico, a demanda por pessoas mais qualificadas será maior; a realização de estudo aprofundado para compreender melhor as possíveis implicações da mudança na participação do público masculino; a averiguação da necessidade de ação específica endereçada para a faixa mais jovem dos responsáveis familiares (abaixo dos 21 anos de idade); e, por fim, o redesenho de políticas assistenciais de modo a incluir alguma medida voltada ao público situado logo acima da linha de vulnerabilidade, e que passou a apresentar necessidades básicas não supridas.

Dentre os resultados relevantes, verificou-se que a idade média das pessoas cadastradas caiu de 46 (antes da pandemia) para 42 (depois do início pandemia). O gráfico 1 mostra como a proporção de pessoas muito jovens (abaixo dos 21 anos) aumentou cinco vezes, de 0,6% para 3,3%:

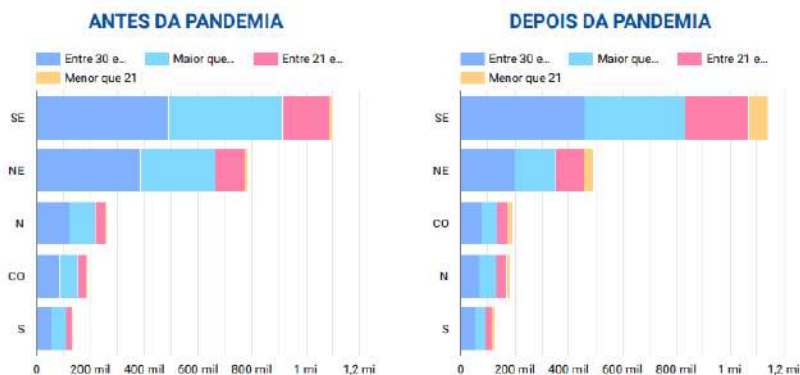
Gráfico 1 | Distribuição das pessoas do CadÚnico por idade



Fonte: Elaborado pelos autores.

A inclusão de novos cadastros com pessoas mais jovens foi maior especialmente no Nordeste do Brasil, conforme gráfico a seguir:

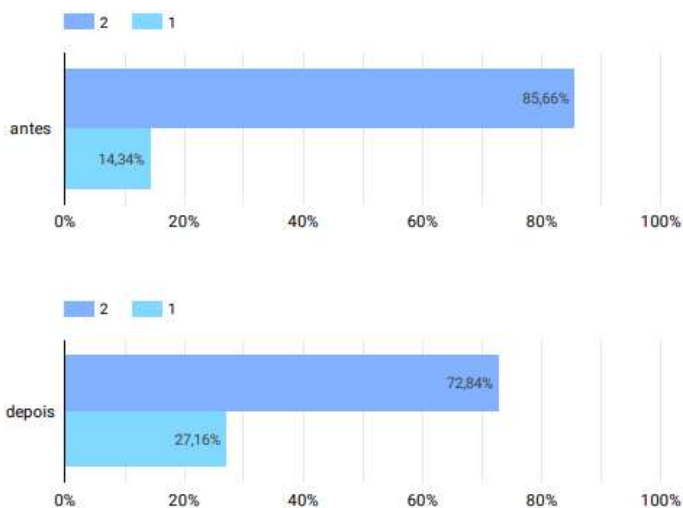
Gráfico 2 | Distribuição das pessoas do CadÚnico por idade e por Região



Fonte: Elaborado pelos autores.

Verificou-se também que houve grande migração de pessoas do sexo masculino para o CadÚnico após a pandemia, visto que a proporção de homens como responsável familiar quase dobrou, passando de 14% para 27%:

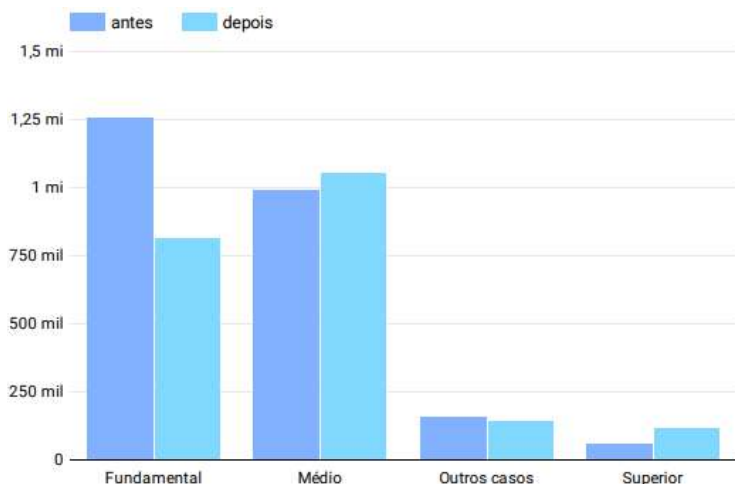
Gráfico 3 | Distribuição do CadÚnico por sexo (Homens = 1, Mulheres = 2)



Fonte: Elaborado pelo autor.

Outra mudança significativa no CadÚnico foi o número de pessoas com nível superior, que, dentro do recorte especificado, dobrou na pandemia, conforme mostrado no gráfico a seguir:

Gráfico 4 | Número de pessoas no CadÚnico por grau de instrução



Fonte: Elaborado pelo autor.

CONCLUSÃO

Foi possível observar mudanças consideráveis no perfil dos inscritos no CadÚnico nos municípios com mais de 500 mil habitantes, destacadamente o rejuvenescimento da população, o aumento da participação masculina e de pessoas com grau de instrução mais elevado. O cadastro mostrou-se peça-chave para transformações sociais e, caso seja devidamente explorado, com o apoio da ciência de dados, pode resolver os problemas da integração e retroalimentação apontados por Direito e Koga (2022). Assim, a resposta à questão de pesquisa é parcial, ou seja, o CadÚnico, por si só, não é suficiente para a revelação plena do perfil das famílias, depende, porém de devida integração, retroalimentação e análise adequada dos dados ali contidos.

Por outro lado, diante das mudanças no perfil aqui observadas, torna-se necessário analisar essas mudanças, no sentido de compreender quais as suas possíveis implicações econômicas e sociais e respondê-las por meio de ações específicas e ágeis, na forma lecionada por Botelho *et al.* (2020). Essas análises devem ser feitas em pesquisas posteriores, complementares e mais aprofundadas.

Nessa linha, a título de aprimoramento da ferramenta, sugere-se a realização dos seguintes estudos por parte dos formuladores da política pública:

- a. Criar um programa específico para a requalificação e atualização de pessoas com nível superior, partindo da premissa de que, num primeiro momento de retomada do crescimento econômico, a demanda por pessoas mais qualificado será maior.
- b. Realizar estudo aprofundado para compreender melhor a mudança na participação do público masculino, que praticamente dobrou, para a amostra analisada, durante a pandemia.
- c. Analisar mais detalhadamente a composição da faixa de Responsáveis Familiares abaixo dos 21 anos e averiguar se há necessidade de ação específica, tendo em vista o aumento desse público em quase sete vezes após a pandemia.

- d. Redesenhar políticas assistenciais de modo a incluir alguma medida voltada ao público situado logo acima da linha de vulnerabilidade e que passou a apresentar necessidades básicas não supridas.

REFERÊNCIAS

BOTELHO, V. *et al.* **Programa de Responsabilidade Social**: Diagnóstico e Proposta – *CDPP*. Disponível em: <<https://cdpp.org.br/2020/09/18/programa-de-responsabilidade-social-diagnostico-e-proposta-2/>>. Acesso em: 26 set. 2022.

CALDAS, R. de M.; SAMPAIO, Y. de S. B. Pobreza no nordeste brasileiro: uma análise multidimensional. **Revista de Economia Contemporânea**, v. 19, n. 1, p. 74–96, 1 jan. 2015. Disponível em: <<http://www.scielo.br/j/rec/a/T7fzd-8My98VMWxw4sPH8rCq/?lang=pt>>. Acesso em: 26 set. 2022.

DIREITO, D. do C.; KOGA, N. M. Instrumentos e integração de políticas públicas: a rede do Cadastro Único. **Revista de Administração Pública**, v. 54, n. 5, p. 1286–1306, 2 nov. 2020. Disponível em: <<http://www.scielo.br/j/rap/a/hN7JX7kx6gFLSN38NZmwtDn/?lang=pt>>. Acesso em: 26 set. 2022.

FIGUEIREDO, E. Programa Auxílio Brasil: ampliação da cobertura, reajuste do valor do benefício e as suas interações com o mercado de trabalho formal. **Nota da Presidência**, n. 10. Brasília: IPEA, 2022.

JOHNS HOPKINS UNIVERSITY & MEDICINE. COVID-19 Dashboard. **Coronavirus Resource Center**, 2022. Disponível em: <https://coronavirus.jhu.edu/map.html>. Acesso em: 3 out. 2022.

MALTA, D. C. *et al.* A pandemia da COVID-19 e as mudanças no estilo de vida dos brasileiros adultos: um estudo transversal, 2020. **Epidemiol. Serv. Saúde**, Brasília, v. 29, n. 4, e2020407, set. 2020. Disponível em <http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S1679-49742020000400025-&lng=pt&nrm-iso>. Acesso em: 3 out. 2022.

MARINHO, E.; LINHARES, F.; CAMPELO, G. Os programas de transferência de renda do governo impactam a pobreza no Brasil?. **Revista Brasileira de Economia**, v. 65, n. 3, p. 267-288, 2011.

MATTA, G.C. *et al.* **Os impactos sociais da Covid-19 no Brasil**: populações vulnerabilizadas e respostas à pandemia [online]. Rio de Janeiro: Observatório Covid 19; Editora FIOCRUZ, 2021, 221 p. Informação para ação na Covid-19 series. ISBN: 978-65-5708-032-0. <https://doi.org/10.7476/9786557080320>.

MINISTÉRIO DA CIDADANIA. **CadÚnico**: o que é e para que serve. Disponível em: <https://www.gov.br/cidadania/pt-br/acoes-e-programas/cadastro-unico/o-que-e-e-para-que-serve-1>. Acesso em: 2 out. 2022.

NERI, Marcelo C. **Mapa da Nova Pobreza**. Rio de Janeiro, RJ – junho/2022 – FGV Social. Disponível em: <https://cps.fgv.br/MapaNovaPobreza>. Acesso em: 1 out. 2022.

OSORIO, R. G. *et al.* **Perfil da pobreza no Brasil e sua evolução no período 2004-2009**. Texto para discussão, 2011.

PAIVA, L.H.; SOUZA, P.H.G.F; BARTHOLO, L; SOARES, S. Evitando a pandemia da pobreza: possibilidades para o programa Bolsa Família e para o Cadastro Único em resposta à COVID-19. **Revista Administração Pública**, 54(4), Jul-Ago 2020. Disponível em: <https://www.scielo.br/j/rap/a/GtG37gSZy4CVCBvdcPcr-gYb/?format=html&lang=pt>. Acesso em: 22 set. 2022

PRÓSPERI, L. de O. **O Programa Bolsa Família, o “efeito preguiça” e o mercado de trabalho**. 2015. 93 f. Dissertação (Mestrado) – Curso de Gestão Pública e Sociedade, Universidade Federal de Alfenas, Varginha, 2015.

ROCHA, S. **Pobreza no Brasil: afinal, de que se trata?** Third ed. [S.l.]: FGV Editora, 2011. Disponível em: https://edisciplinas.usp.br/pluginfile.php/278399/mod_resource/content/1/3649_001.pdf. Acesso em: 26 set. 2022.

ROSSI, C. **Os impactos dos programas condicionais de transferência de renda na oferta de trabalho dos jovens nem-nem**. 2017. Dissertação (Mestrado em Economia Aplicada) - Escola Superior de Agricultura Luiz de

Queiroz, Universidade de São Paulo, Piracicaba, 2017. doi:10.11606/D.11.2017.tde-15082017-182203. Acesso em: 2022-07-20.

SOUZA, T. A. A. de; COUTO, A. C. L. Programa Bolsa Família: características e perfil da pobreza entre os beneficiários nos estados do Paraná e Bahia. **Informe Gepec**, v. 19, n. 2, p. 41-56, 2015.

TOMAZINI, C. Beyond consensus: ideas and advocacy coalitions around cash transfer programs in Brazil and Mexico. **Critical Policy Studies**, p. 23-42, 2017. Disponível em: <<https://bit.ly/3H2r1MG>>. Acesso em: 20 jul. 2022.



**TEMA: SEGURANÇA
ALIMENTAR E NUTRICIONAL**

FERRAMENTA: APRENDIZADO DE
MÁQUINA SUPERVISIONADO

PANORAMA DA SITUAÇÃO DE INSEGURANÇA ALIMENTAR E NUTRICIONAL DOS MUNICÍPIOS BRASILEIROS APÓS A PANDEMIA DA COVID-19¹

ELIAS DE OLIVEIRA, ERICA RAMALHO DE OLIVEIRA, GUSTAVO DE QUEIROZ CHAVES, ISVALDO LIMA FILHO, RAFAELA DE SÁ GONÇALVES E RONAN SILVA

INTRODUÇÃO

A avaliação da situação de Insegurança Alimentar e Nutricional (INSAN) de um território pode ser realizada a partir da utilização de diferentes indicadores que captem as diversas dimensões relacionadas à Segurança Alimentar e Nutricional (SAN) das populações. Trata-se de uma agenda que requer esforço permanente de diagnóstico das diferentes realidades encontradas no país, de forma a produzir evidências para identificação de grupos vulneráveis, melhorando a focalização e qualificação do processo de formulação, priorização e implementação de políticas públicas.

¹ Este trabalho foi realizado durante o curso, no Laboratório de Casos de Políticas Públicas Sociais, sob orientação dos professores: Jackson De Toni, Natalia Massaco Koga, Helio Bomfim de Macedo Filho e Jeniffer de Paula Chaves.

Destaca-se, nesse contexto, a existência de estudo intitulado Mapeamento da Insegurança Alimentar e Nutricional (MAPAINSAN), desenvolvido e publicado pela Câmara Interministerial de Segurança Alimentar e Nutricional (CAISAN) nos anos de 2016 (ano referência 2014) e 2018 (ano referência 2016) com o objetivo de identificar os territórios mais vulneráveis em relação à situação de INSAN, a partir de dados do Cadastro Único e do Sistema Nacional de Vigilância Alimentar e Nutricional (SISVAN).

Considerando, assim, os critérios utilizados no MAPAINSAN, bem como a importância da produção de informações atualizadas sobre a SAN no Brasil, especialmente após a emergência de saúde pública de importância internacional provocada pelo novo coronavírus (covid-19), propôs-se, com o presente estudo, a identificação de um modelo de classificação do nível de vulnerabilidade municipal em SAN, com base em dados agregados do estado nutricional de crianças beneficiárias do Bolsa Família/Auxílio Brasil e dados agregados das condições socioeconômicas das famílias inscritas no Cadastro Único, para os anos de 2019 e 2021/2022.

Por fim, registra-se que o estudo proposto considerou, exclusivamente, informações constantes em registros administrativos de sistemas do Governo Federal entre 2019 e 2021/2022. Assim, eventuais fragilidades na coleta e na atualização dos dados utilizados na modelagem, em especial àquelas ocasionadas pela emergência da covid-19 (iniciada no ano de 2020), não foram avaliadas no presente estudo. Da mesma forma, não foi possível ponderar eventuais impactos decorrentes da implementação de programas temporários de transferência de renda durante a pandemia da covid-19, programas esses que possuíam alcance mais amplo do que as políticas de transferências de renda atualmente vigentes.

DESENVOLVIMENTO

O conceito de Segurança Alimentar e Nutricional adotado no Brasil — estabelecido pelo art. 3º da Lei n. 11.346, de 15 de setembro de 2006² — é amplo e abarca diversos aspectos do fenômeno, como as dimensões de disponibilidade, regularidade, acesso e consumo de alimentos, além de questões ambientais, culturais, econômicas, regionais e sociais. Nesse sentido, a avaliação da situação de INSAN de indivíduos e territórios pode ser feita sob diferentes perspectivas.

As escalas de percepção, como a Escala Brasileira de Segurança Alimentar e Nutricional (EBIA), são comumente adotadas como medida direta de percepção e vivência de insegurança alimentar e fome no nível domiciliar. Por outro lado, em razão da maior complexidade de aplicação das escalas de percepção, que requerem a realização de pesquisa para aplicação dos questionários, utiliza-se com frequência medidas indiretas de INSAN, como indicadores de saúde, socioeconômicos e demográficos, que são importantes marcadores da situação de pobreza e vulnerabilidade social que normalmente se encontram as famílias em situação de INSAN.

Nesse sentido, o presente estudo, com o objetivo de verificar a situação de INSAN dos municípios brasileiros no cenário de pós-pandemia, de forma tempestiva e não onerosa, desenvolveu, por meio de estratégias de aprendizado de máquina (*Machine Learning*), treinamento supervisionado de modelo de classificação do nível de vulnerabilidade municipal em INSAN, partindo de classificação já existente, criada pelo estudo Mapeamento da Insegurança Alimentar e Nutricional em 2016.

Qifang Bi *et al.* (2019) caracterizam o aprendizado de máquina como sendo uma subárea da Inteligência Artificial que busca a utilização de técnicas que possibilitem que computadores “aprendam” com a “experiência”, baseadas na análise de dados passados (de teste), de modo indutivo, por meio da iden-

2 Art. 3º A segurança alimentar e nutricional consiste na realização do direito de todos ao acesso regular e permanente a alimentos de qualidade, em quantidade suficiente, sem comprometer o acesso a outras necessidades essenciais, tendo como base práticas alimentares promotoras de saúde que respeitem a diversidade cultural e que sejam ambiental, cultural, econômica e socialmente sustentáveis (BRASIL, 2006).

tificação de padrões. Em vez de se especificar uma série de regras (por meio de uma linguagem de programação), a máquina (ou o algoritmo) teria como apreendê-las a partir da observação de um conjunto de dados, identificando padrões que determinem, com certa acurácia, os resultados pretendidos, com aplicações, especialmente, com objetivo de predição e otimização.

O aprendizado supervisionado, utilizado neste trabalho, baseia-se na utilização de um conjunto de dados de treinamento totalmente rotulados, com o resultado (*target*) conhecido para cada observação. Além disso, como a variável de desfecho (*target*) definida foi uma variável categórica, trata-se aqui, portanto, de um problema de classificação e não de regressão, na qual se teria como desfecho uma variável quantitativa contínua (JAMES; WITTEN; HASTIE; TIBSHIRANI, 2013).

Importante enfatizar que o aprendizado de máquina tem experimentado um grande progresso nas últimas duas décadas, despontando como uma tecnologia aplicada amplamente no setor produtivo e em vários ramos da ciência, como Biologia, Cosmologia e Ciências Sociais (JORDAN; MITCHEL, 2015).

METODOLOGIA

MAPEAMENTO DA INSEGURANÇA ALIMENTAR E NUTRICIONAL (MAPAINSAN)

O MAPAINSAN é um estudo desenvolvido e publicado pela CAISAN nos anos de 2016 (dados de 2014/2015) e 2018 (dados de 2016/2017), que tinha como um dos objetivos a identificação dos municípios mais vulneráveis em relação à situação de INSAN de sua população. Os municípios foram classificados em vulnerabilidade média, alta e muito alta.

Para classificação dos municípios, o referido estudo utilizou-se do método estatístico análise de cluster, técnica usada para classificar elementos em grupos, de forma que elementos dentro de um mesmo grupo sejam muito parecidos e elementos em diferentes grupos sejam distintos entre si. As va-

riáveis escolhidas para compor a análise de cluster foram o Déficit de Altura para Idade (DAI) e o Déficit de Peso para Idade (DPI) em crianças, por serem consideradas variáveis proxy da situação de INSAN de uma população em situação de vulnerabilidade social.

Considerando o nível de detalhamento e de representatividade exigido para a avaliação da desnutrição infantil, o estudo optou pela utilização da base de dados nominais do Sistema de Vigilância Alimentar Nutricional (SISVAN/ Sistema de Gestão do Bolsa Família/DATASUS), cuja cobertura é bastante significativa para as crianças menores de 5 anos beneficiárias do Bolsa Família.

Já para a investigação dos determinantes da desnutrição e INSAN, o estudo optou por trabalhar com dados nominais do Cadastro Único, por conter informações consistentes e detalhadas sobre as famílias das crianças analisadas, a saber: renda familiar per capita (nível de pobreza) e características de infraestrutura do domicílio (acesso à água e ao esgotamento sanitário).

Após contratação de consultoria para realização do cruzamento das referidas bases de dados, classificou-se os municípios a partir dos dados de desnutrição infantil (vulnerabilidade média, alta e muito alta) e demonstrou-se o alto grau de correlação desses indicadores com as condições socioeconômicas das famílias em que vivem tais crianças.

MODELO DE APRENDIZADO DE MÁQUINA

A base de dados gerada pelo MAPAINSAN, com a classificação de cada um dos 5.570 municípios segundo nível de vulnerabilidade em INSAN, serviu como base de treinamento para o modelo do presente trabalho. O modelo foi então utilizado para classificar o nível de vulnerabilidade dos municípios nos anos de 2019 e 2022 a partir de dados agregados disponibilizados publicamente em SISVAN e VIS DATA 3 beta (dados do Cadastro Único).

Para o ano de 2019, utilizou-se dados do SISVAN - 2019 e do Cadastro Único - dez/2019. Para o ano de 2022, utilizou-se dados do SISVAN - 2021 e do Cadastro Único - mar/2022.

O modelo utilizou como variáveis independentes:

- “DPI”: Déficit de Peso para Idade — proporção de crianças menores de 5 anos, acompanhadas nas condicionalidades de saúde do Programa Bolsa Família/Auxílio Brasil, que apresentam peso corporal abaixo do limite de normalidade aceitável para a idade, o que está associado à perda de peso recente e/ou à deficiência de estatura (fonte: SISVAN/Sistema de Gestão do Bolsa Família).
- “DAI”: Déficit de Altura para Idade — proporção de crianças menores de 5 anos, acompanhadas nas condicionalidades de saúde do Programa Bolsa Família/Auxílio Brasil, que apresentam estatura abaixo do limite de normalidade aceitável para a idade, caracterizando um déficit crônico (fonte: SISVAN/Sistema de Gestão do Bolsa Família).
- “sem_agua”: percentual de famílias do Cadastro Único que não possuem acesso adequado à água, considerando para isso domicílios que não possuem acesso à rede geral, poço ou nascente, ou cisternas.
- “sem_esgoto”: percentual de famílias do Cadastro Único que não possuem esgoto adequado, considerando para isso domicílios que não possuem acesso a esgotamento sanitário por rede coletora ou fossa séptica.
- “pobreza”: percentual de famílias do Cadastro Único que vivem em situação de pobreza e extrema pobreza.

E como variável dependente (*target*):

- Nível de vulnerabilidade em SAN (4 classes):
 1. muito alta e alta³; 2. média e 3. baixa⁴

A definição do modelo priorizou métodos em que o erro de classificação fosse conservador, ou seja, sendo preferível a classificação como mais vul-

3 Para melhor desempenho, o modelo agregou os níveis de vulnerabilidade alta e muito alta do MAPAINSAN em um só nível.

4 A classificação “baixa” foi imputada no modelo para todos os municípios que não entraram no MAPAINSAN, que incluiu apenas os municípios com média igual ou superior a 10,1% de DAI.

nerável de uma localidade menos crítica do que o cenário contrário (nesse caso, a medida de sensibilidade — *recall* — do modelo é mais relevante, pois os falsos negativos são mais indesejáveis do que os falsos positivos).

Fez-se a escolha do algoritmo por meio do pacote “Lazy Predict”, que tem o objetivo de auxiliar a construir e testar uma série de modelos básicos (sem “tunning” dos hiperparâmetros) para vários algoritmos (estimadores). O modelo escolhido, Random Forest, teve desempenho da ordem de 90% em suas métricas de avaliação — acurácia, precisão, sensibilidade e média harmônica das duas anteriores (F1-score).

No modelo, a importância das variáveis se dividiu nas proporções de 65,8% para o Déficit de Altura (DAI), 17% para o Déficit de Peso (DPI), e 5,6%, 5,8% e 5,8% para as variáveis do Cadastro Único, respectivamente, de acesso inadequado à água, sistema inadequado de esgotamento sanitário e famílias em situação de pobreza.

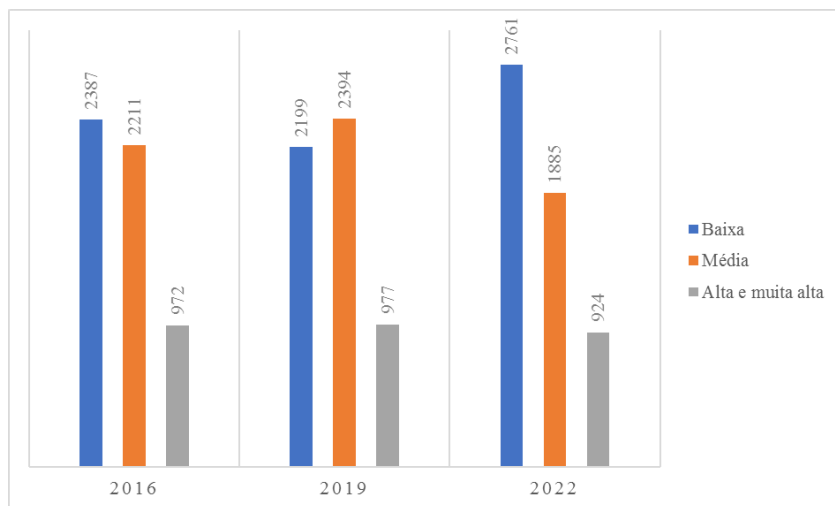
A premissa das avaliações, utilizando esse modelo, contempla a hipótese de que, desde 2016, a vulnerabilidade alimentar tenha se mantido adequadamente representada por esses 5 fatores em razões similares às da modelagem.

RESULTADOS

Nesta seção serão apresentados alguns achados em relação à evolução da situação de INSAN no Brasil, grandes regiões e unidades da federação, a partir da aplicação do modelo.

Conforme apresentado no Gráfico 1, entre a linha de base dos dados (2016) e o ano anterior à pandemia (2019) houve uma ampliação dos municípios classificados como em média e alta/muito alta vulnerabilidade. Por outro lado, entre os anos de 2019 e 2022, há uma considerável ampliação dos municípios em baixa vulnerabilidade, um aumento observado de 10% a nível nacional.

Gráfico 1 | Número de municípios do Brasil por nível de vulnerabilidade em 2016, 2019 e 2022.



Fonte: Elaborado pelos autores (2022).

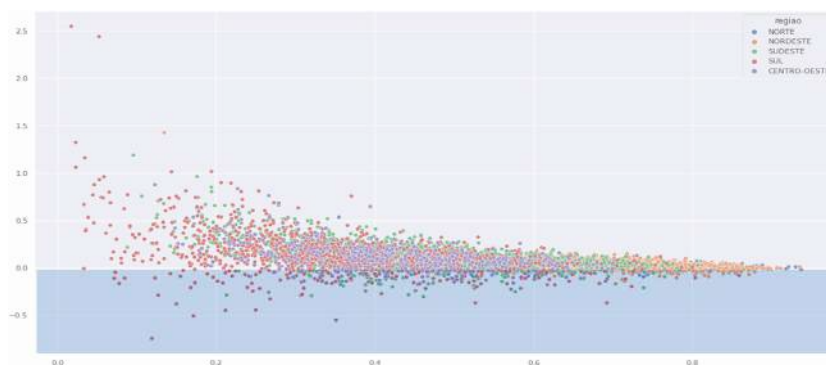
Como os indicadores de DAI e DPI advindos do SISVAN são os que mais têm peso no modelo, um dos pontos que pode ter influenciado tal resultado é a diminuição do número de crianças acompanhadas pelas condicionalidades de saúde do Programa Bolsa Família durante a pandemia, em razão da suspensão de procedimentos operacionais e de gestão do Programa. Com base nos registros do SISVAN, o número de crianças menores de 5 anos acompanhadas pelas condicionalidades de saúde do Programa Bolsa Família caiu aproximadamente 40% entre os anos de 2019 e 2021. Em 2019 foram 3.684.281, contra 2.258.708 crianças em 2021.

Por outro lado, em contraposição ao cenário de melhora na classificação de INSAN, o percentual de famílias do Cadastro Único que vivem abaixo da linha da pobreza passou de 57% em 2019 para 61% em 2022, um aumento estatisticamente significativo.

Sobre a área cinza na Figura 1 são mostrados os municípios onde a diferença de percentual de famílias em situação de pobreza aumentou após a pandemia. Abaixo, na área azul, localizam-se os municípios em que esse

índice apresentou melhora, isto é, em que o percentual de famílias pobres sofreu redução quando comparado a 2019.

Figura 1 | Dispersão dos municípios em função do aumento ou redução do percentual de famílias em situação de pobreza, por região



Fonte: Elaborado pelos autores (2022)

RESULTADOS REGIONAIS

Em 2016, o percentual de crianças acompanhadas com DAI era maior nos municípios das regiões Norte e Nordeste, enquanto a variabilidade do déficit era mais significativa entre os municípios do Norte e Sul do país.

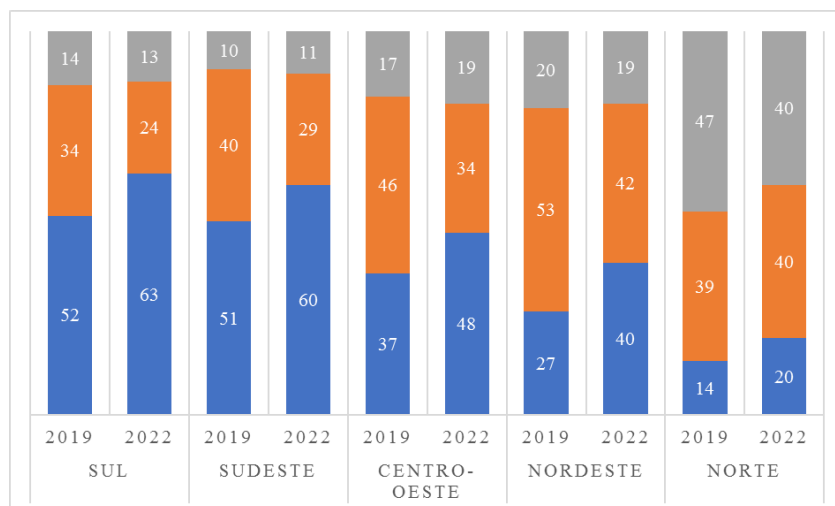
As regiões Sul e Sudeste são as que apresentam maior similaridade na distribuição de seus municípios quanto à vulnerabilidade alimentar estimada.

Nas cinco regiões, houve um aumento da quantidade de municípios classificados como de baixa vulnerabilidade alimentar entre os anos de 2019 e 2022.

A maior queda do número de municípios em vulnerabilidade alta e muito alta (cor cinza no gráfico) se deu na região Norte (7%), enquanto o maior aumento do número de municípios em vulnerabilidade baixa (cor azul no gráfico) aconteceu na região Nordeste (13%).

À exceção da região Norte, as variações de classes de vulnerabilidade alimentar concentraram-se nos níveis baixos e médio.

Gráfico 2 | Percentual de municípios de cada região por níveis crescentes de vulnerabilidade em 2019 e 2022.



Fonte: Elaborado pelos autores (2022).

Embora as regiões Sul e Sudeste tenham apresentado a maior variação de localidades em que o percentual de famílias vivendo na pobreza cresceu, o Nordeste é a região onde os municípios concentram-se em índices médios mais altos.

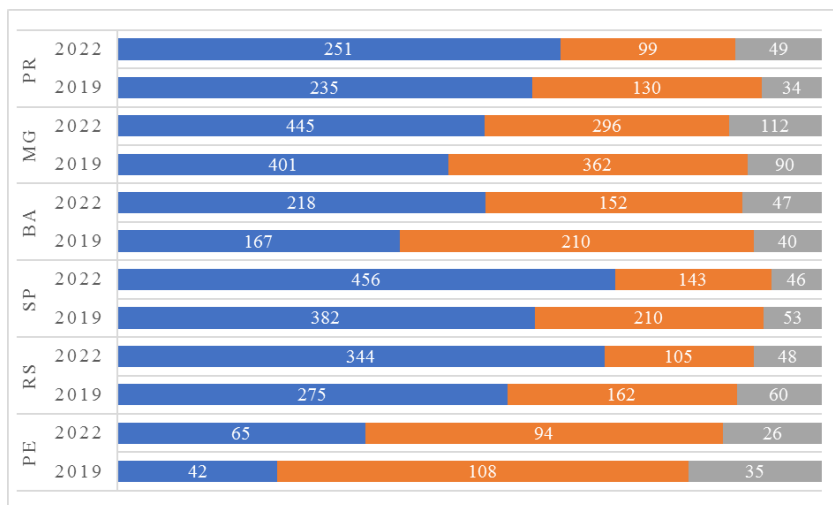
RESULTADOS ESTADUAIS

As Unidades da Federação, de forma geral, não apresentaram comportamento homogêneo quanto à alteração das classes de vulnerabilidade alimentar previstas para 2019 e 2022.

Alguns estados, como Pernambuco, Rio Grande do Sul e São Paulo, por exemplo, tiveram melhoria representada pela redução da quantidade de municípios em classes alta (cor cinza) e média vulnerabilidade (cor laranja).

Outras unidades, como Bahia, Minas Gerais e Paraná, seguiram a tendência nacional e regional de ampliação da quantidade de municípios em baixa vulnerabilidade alimentar (cor azul), mas também tiveram sua quantidade de municípios em alta vulnerabilidade aumentada em 2022.

Gráfico 3 | Número de municípios em UFs específicas por níveis crescentes de vulnerabilidade em 2019 e 2022



Fonte: Elaborado pelos autores (2022).

Quanto ao percentual de famílias abaixo da linha de pobreza, o município com maior destaque no período foi São Francisco do Sul/RS, cuja redução foi de 74% em 2022 comparado a 2019, seguido de Mirante da Serra/RO (55%) e Rio do Oeste/SC (51%).

Sete dos dez municípios com maior aumento no percentual de famílias abaixo da linha de pobreza estão localizados na região Sul do Brasil. Esse grupo possuía percentuais de famílias abaixo da linha de pobreza antes da pandemia em índices menores que 20%.

Quadro 1 | Municípios com maior aumento no percentual de famílias abaixo da linha de pobreza.

Município/UF	Aumento de famílias abaixo da linha de pobreza
São José do Inhacorá/RS	254,5%
Lacerdópolis/SC	243,6%
Fernando de Noronha/PE	142,4%
Vespasiano Correa/RS	132,2%
Serra da Saudade/MG	118,8%
Anta Gorda/RS	116,0%
Três Arroios/RS	106,0%
Quatro Barras/PR	101,6%
Nova Petrópolis/RS	101,3%
Bady Bassitt/SP	96,2%

Fonte: Elaborado pelos autores (2022).

CONCLUSÃO

De modo geral, pode-se concluir, pela aplicação do modelo, que os municípios apresentaram melhora nos níveis de INSAN entre 2019 e 2022. Enquanto, em 2019, cerca de 40% dos municípios encontravam-se em níveis baixos de vulnerabilidade, esse percentual subiu para 50% dos municípios em 2022.

Apenas 975 municípios pioraram de situação entre os anos de 2019 e 2022, passando de um nível de vulnerabilidade para outro de maior criticidade. Observa-se que o fator preponderante de variação nos municípios que pioraram seu nível de vulnerabilidade foi o indicador Déficit de Altura para Idade (DAI).

Quanto aos demais indicadores, observa-se uma piora nos níveis de acesso a esgotamento sanitário nas regiões Norte e Nordeste, enquanto nas regiões Sul, Sudeste e Centro-Oeste, tal variação se deu em relação ao nível de pobreza, que se apresentou mais elevado. Não foi possível, no entanto, identificar uma correlação entre a piora desses indicadores com mudanças nos níveis de INSAN.

Considerando que os níveis médios de vulnerabilidade da região Nordeste já eram elevados, a pandemia não trouxe grande variação a esses municípios. Uma possível explicação para esse cenário pode ser a não descontinuidade dos auxílios de transferência de renda que já abrangiam grande parte dessas famílias em situação de extrema pobreza.

Ademais, pondera-se que, em razão da multifatorialidade da insegurança alimentar e nutricional de um território, nenhum indicador isolado é capaz de proporcionar um diagnóstico completo da situação. Nesse sentido, faz-se necessária a utilização de indicadores complementares para uma avaliação mais fidedigna da realidade.

Por fim, entende-se como importante a realização de uma nova rodada de avaliação do modelo com dados de 2022 e/ou 2023, de forma (1) a mitigar eventuais impactos relacionados a problemas na coleta e na atualização de dados das famílias nos registros administrativos do Governo Federal, especialmente devido às dificuldades impostas pela covid-19; e; (2) a ponderar eventuais impactos decorrentes dos programas temporários de transferência de renda que estiveram vigentes, principalmente, entre 2020 e 2021.

REFERÊNCIAS

BRASIL. Lei n. 11.346, de 15 de setembro de 2006. **Cria o Sistema Nacional de Segurança Alimentar e Nutricional com vistas em assegurar o direito humano à alimentação adequada e dá outras providências.** Diário Oficial da União 2006; 18 set.

BRASIL. Câmara Interministerial de Segurança Alimentar e Nutricional. **Mapeamento da Insegurança Alimentar e Nutricional com foco na Desnutrição a partir da análise do Cadastro Único e do SISVAN 2014**. Brasília, DF, out. 2016. Disponível em: http://aplicacoes.mds.gov.br/sagirmsps/noticias/arquivos/files/Estudo_T%C3%A9cnico_MAPAINSAN_2014.pdf. Acesso em: 1 out. 2022.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning: with Applications in R**. Springer, 2013. Disponível em: https://www.academia.edu/43297642/Introduction_to_Statistical_Learning?from=cover_page. Acesso em: 1 out. 2022.

JORDAN, M. I; MITCHEL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**. v. 349, p. 255-260, jul. 2015. DOI: 10.1126/science.aaa8415. Disponível em: <http://www.cs.cmu.edu/~tom/pubs/Science-ML-2015.pdf>. Acesso em: 20 ago. 2022.

QIFANG BI. *et al.* What is Machine Learning? A Primer for the Epidemiologist, **American Journal of Epidemiology**, v. 188, n. 12, p. 2222-2239, dez. 2019. Disponível em: <https://doi.org/10.1093/aje/kwz189>. Acesso em: 20 ago. 2022.

SOBRE OS AUTORES, ORGANIZADORES E COLABORADORES

(1) **Rodrigo Torres** — Diretor de Educação Executiva da Enap e membro da carreira de Especialista em Políticas Públicas e Gestão Governamental. Possui Mestrado e Bacharelado em Relações Internacionais. Foi Secretário Nacional de Promoção dos Direitos da Criança e do Adolescente e presidente do Conselho Nacional dos Direitos da Criança e do Adolescente (CONANDA). Na Enap já atuou como Chefe de Gabinete.

(2) **Natália Massaco Koga** — Doutora em Ciência Política pela University of Westminster/Reino Unido, mestre em Administração Pública e Governo pela FGV-EAESP e Graduada em Direito pela Faculdade de Direito da Universidade de São Paulo. Membro da carreira de Especialista em Políticas Públicas e Gestão Governamental do Ministério da Economia. Atualmente, pesquisadora no Instituto de Pesquisa Econômica Aplicada (Ipea) e professora do Mestrado Profissional em Governança e Desenvolvimento da Escola Nacional de Administração Pública (Enap) nas áreas de políticas públicas e uso de evidências, capacidades estatais e democracia. Teve experiência profissional como gestora federal no Ministério do Desenvolvimento Social (MDS), Ministério do Planejamento e Orçamento e Ministério de Minas e Energia (MME).

(3) **Alex Lopes Pereira** — Doutor em Engenharia Eletrônica e Computação pelo Instituto Tecnológico de Aeronáutica (ITA). É integrante da carreira de Analista de Planejamento e Orçamento. Atualmente, trabalha no Ministério da Economia. Foi Analista de Ciência e Tecnologia do Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM) e Empreendedor em *startups* de tecnologia da informação.

(4) **Hélio Bomfim de Macedo Filho** — Bacharel em Ciência da Computação pela Universidade Federal do Ceará (UFC), com distinção acadêmica *Magna Cum Laude*, em 2007. Mestre e Doutor em Engenharia de Sistemas e Computação

pela Universidade Federal do Rio de Janeiro (UFRJ). É autor de artigos publicados em periódicos internacionais, como o *Theoretical Computer Science*, *Discrete Applied Mathematics* e *Algorithmica*. Lidera os projetos da área de dispositivos médicos na Anvisa.

Jackson De Toni — Economista, Mestre em Planejamento Regional e Urbano pela Universidade Federal do Rio Grande do Sul (UFRGS), Doutor em Ciência Política pela Universidade de Brasília (UnB) e Especialista em Ciência de Dados aplicada a Políticas Públicas pela Escola Nacional de Administração Pública (Enap). Analista de Produtividade e Inovação na Agência Brasileira de Desenvolvimento Industrial (ABDI). Professor de Políticas Públicas e Planejamento Governamental na Enap, IBMEC-DF, IESB e FGV-DF.

Rachel Pereira Dorneles — Atua na Enap como coordenadora do curso de Especialização. Tem mestrado em Educação, Especializações em Políticas Públicas de Educação, em Inovação e Tecnologias na Educação e em Gestão Pública. Trabalhou no antigo Ministério do Desenvolvimento Social nas parcerias com programas de Educação. É servidora do Ministério da Educação.

DISCENTES

(5) **Alexandre Martins Nogueira** — Bacharel em Ciência da Computação e pós-graduado em Administração de Sistemas. Atua há 17 anos na Controladoria-Geral da União, atualmente é Auditor especialista em tecnologia da informação e um entusiasta na área de Ciência de Dados.

(6) **Gustavo Queiroz Chaves** — Auditor Federal de Finanças e Controle pela Controladoria-Geral da União (CGU), Mestre em Ciências, em Engenharia Metalúrgica e de Materiais (Coppe/UFRJ), Certificado em *Risk Management Assurance* (IIA), Especialista em Ciência de Dados aplicada a Políticas Públicas pela Escola Nacional de Administração Pública (Enap) e Bacharel em Engenharia Metalúrgica (IME).

(7) **Isvaldo Silveira Lima Filho** — Mestre em Administração Pública pela Universidade Federal Rural de Pernambuco. Bacharel em Ciências da Computação pela Universidade Federal de Pernambuco. Auditor Federal de Finanças e Controle e Supervisor de Auditoria na Controladoria-Geral da União (CGU).

(8) **José Leonardo Ribeiro Nascimento** — Auditor Federal de Finanças e Controle na Controladoria-Geral da União (CGU) há vinte anos, possui graduação em Ciências Contábeis pela Universidade Federal de Sergipe (UFS) e especialização em Auditoria Governamental e Contabilidade Pública pela Faculdade de Administração e Negócios de Sergipe (Fanese). Em 2012, participou do *Minerva Program na George Washington University* nos EUA e atualmente trabalha com Ciência de Dados aplicada na avaliação de políticas públicas.

(9) **Matheus Scatolino de Rezende** — Auditor Federal de Finanças e Controle desde 2012, graduado em Ciência da Computação pela PUC-MG e Administração pela UFMG, especialista em Ciência de Dados aplicado em Políticas Públicas pela ENAP. Atualmente, exerce o cargo de chefe de divisão de auditoria de TI na CGU.

(10) **Mônica Rondina** — Engenheira civil, arquiteta e urbanista. Trabalha como Auditora Federal de finanças e controle da Controladoria Geral da União. Atualmente audita empresas estatais dos setores de energia e petróleo.

(11) **Rafael Leal Garcia** — Formado em Ciência da Computação pela Universidade Federal da Bahia. Trabalha na CGU desde 2007.

(12) **Rodrigo Santos de Paula** — Auditor Federal de Finanças e Controle da Controladoria-Geral da União (CGU) desde 2006, Auditor-chefe na ANP em 2019 e 2020, MBA em Gestão Pública, graduação em Análise de Sistema na Universidade do Estado do Rio de Janeiro (UERJ).

(13) **Thiago Morais Furuchima** — Especialista em Ciência de Dados pela Universidade de São Paulo (USP), Bacharel em Ciência da Computação pela Universidade Estadual do Oeste do Paraná (Unioeste).

Ciro Jonatas de Souza Oliveira — Especialista em Ciência de Dados aplicada a Políticas Públicas pela Escola Nacional de Administração Pública (Enap) e Mestre em Administração pela Universidade Federal da Bahia. Auditor Federal de Finanças e Controle da Controladoria-Geral da União (CGU).

Elias Fernandes de Oliveira — Auditor Federal de Finanças na CGU. Formado em Ciências Biológicas pela Universidade de Brasília e Pós-Graduado em Ciência de Dados pela ENAP. Atua com auditoria governamental nas áreas de Previdência e Assistência Social.

Érica Ramalho de Oliveira — Gerente técnica de gerenciamento de risco na Agência Nacional de Aviação Civil - ANAC.

Glauber Volkmer — Mestre em Administração pela Universidade Federal da Bahia (UFBA). Especialista em Controle Público pelo Instituto de Contas do Tribunal de Contas do Estado de Santa Catarina (ICON TCE/SC) e Criptografia e Segurança em Redes pela Universidade Federal Fluminense (UFF). Bacharel em Ciências da Computação pela Universidade Federal de Santa Catarina (UFSC). Auditor Federal de Finanças e Controle da Controladoria-Geral da União desde 2006.

Marcos Cândido de Paula Rezende — Trabalha com o tema de educação superior e tecnológica na Controladoria-Geral da União.

Rafaela Gonçalves — Graduada em Ciência Política pela Universidade de Brasília (UnB) e Especialista em Ciência de Dados aplicada a Políticas Públicas pela Escola Nacional de Administração Pública (Enap). É integrante da carreira de Especialista em Políticas Públicas e Gestão Governamental desde 2009. Atualmente exerce suas atividades na Secretaria de Avaliação e Gestão da Informação do Ministério da Cidadania.

Romualdo Anselmo dos Santos — Auditor Federal de finanças e controle na Controladoria-Geral da União (CGU) e professor substituto na Faculdade de Ciências Contábeis da Universidade Federal da Bahia. Mestre e Doutor em Ciência Política, atua nas áreas de controle interno, prevenção e combate à corrupção e políticas públicas.

Ronan Luiz da Silva — Graduado em Ciência da Computação, Especialista em Gerenciamento de Projetos pela Fundação Getúlio Vargas (FGV). Analista de Planejamento e Orçamento no governo federal desde 2012. Mestrado em Desenvolvimento e Políticas Públicas pelo Instituto de Pesquisa Econômica Aplicada (Ipea). Especialização em Ciência de Dados aplicada a Políticas Públicas pela Escola Nacional de Administração Pública (Enap).

Sérgio Carvalho — Servidor do Ministério da Agricultura, Pecuária e Abastecimento - MAPA. Graduado em Administração de Empresas e Especialista em Gestão Pública.

DEPOIMENTOS DE ALUNOS

“O curso de especialização foi muito bem conduzido pela equipe da ENAP e pelos professores convidados, com foco em temática cada vez mais relevante no âmbito de avaliação de políticas públicas. Destaco, também, a excelente rede de colegas que se forma ao final do curso.”

Alexandre Augusto Silva

“O curso é fantástico, é muito bem organizado pela ENAP, sendo ministrado por professores que demonstram proficiência e experiência na matéria. Apesar de já possuir alguma experiência na área de tecnologia, as aulas práticas e os exercícios permitiram solidificar conhecimentos em ciência de dados.”

Alexandre Martins Nogueira

“A especialização nos deu um leque grande de conhecimentos que poderemos explorar e potencializar durante toda a vida profissional e acadêmica. Apreendi sobre muitas ferramentas e tecnologias que irei aplicar e explorar de forma prática. A qualidade do corpo docente, equipe de apoio e material didático foi muito alta.”

Ciro Jonatas Oliveira

“A especialização em Ciência de Dados Aplicada a Políticas Públicas proporcionou o meu aperfeiçoamento na aplicação de técnicas e tecnologias de análise de dados para subsidiar as tomadas de decisões na Administração Pública.”

Ronaldo Gonçalves Borges

Este livro aborda uma atividade de formação ímpar. Sua singularidade repousa no escopo atualíssimo dos temas, a ciência de dados e o campo das políticas públicas. Outra dimensão que o singulariza é relatar trabalhos de ciência de dados aplicada. A experiência formativa, assim, não significa apenas um campo de curadoria de temas, de transmissão de conhecimentos ou recepção passiva de conteúdos. Os alunos, convertidos em protagonistas do processo, assumem a condição de analistas de políticas concretas e reais. Os trabalhos demonstram a conexão entre o saber acumulado e o debate de propósito, da pertinência e das evidências em políticas públicas. Essa construção é o centro de uma teoria de mudança, condição necessária para o sucesso da ação governamental.

Jackson De Toni

ENAP

ISBN 978-65-87791-30-2



9 786587 791302