



Universidade Federal da Paraíba - UFPB
Centro de Ciências Sociais Aplicadas
Programa de Pós-Graduação em Economia do Setor Público

**Avaliação de risco de irregularidades em
processos de aposentadoria no âmbito do
Tribunal de Contas do Estado da Paraíba: uma
aplicação baseada em aprendizagem de máquina
supervisionada**

Pedro de Souza Fleury

João Pessoa-PB, Novembro de 2024

Pedro de Souza Fleury

Avaliação de risco de irregularidades em processos de aposentadoria no âmbito do Tribunal de Contas do Estado da Paraíba: uma aplicação baseada em aprendizagem de máquina supervisionada

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Economia do Setor Público da UFPB, como parte dos requisitos necessários para a obtenção do Título de Mestre em Economia do Setor Público.

Universidade Federal da Paraíba – UFPB

Centro de Ciências Sociais Aplicadas - CCSA

Programa de Pós-Graduação em Economia do Setor Público

Orientador: Hilton Martins de Brito Ramalho

João Pessoa-PB

Novembro de 2024

Catálogo na publicação
Seção de Catalogação e Classificação

F618a Fleury, Pedro de Souza.

Avaliação de risco de irregularidades em processos de aposentadoria no âmbito do Tribunal de Contas do Estado da Paraíba : uma aplicação baseada em aprendizagem de máquina supervisionada / Pedro de Souza Fleury. - João Pessoa, 2024.

76 f. : il.

Orientação: Hilton Martins de Brito Ramalho.
Dissertação (Mestrado) - UFPB/CCSA.

1. Aposentadoria - Setor público. 2. Auditoria - Processos de aposentadoria. 3. Risco de irregularidades - Avaliação. 4. Aprendizagem supervisionada - Máquina. 5. Aposentadoria - Modelo de avaliação - Eficiência. I. Ramalho, Hilton Martins de Brito. II. Título.

UFPB/BC

CDU 35.082.8(043)



Universidade Federal da Paraíba
Centro de Ciências Sociais Aplicadas
Programa de Pós-Graduação em Economia do Setor Público

Ata da reunião da Banca Examinadora designada para avaliar o trabalho de dissertação do mestrando **Pedro de Souza Fleury** submetido para obtenção do grau de mestre em Economia área de concentração em **Economia do Setor Público**.

Aos 27 dias do mês de novembro do ano dois mil e vinte e quatro, às 09 horas, na sala virtual realizada por meio do link <https://meet.google.com/upn-hrri-dap>, reuniram-se em cerimônia pública os membros da Banca Examinadora, constituída pelos professores doutores **Hilton Martins de Brito Ramalho** (Orientador), **Aléssio Tony Cavalcanti de Almeida** (Examinador Interno) e **José Jorge Dias Lima Júnior** (Examinador Externo), a fim de examinarem o candidato ao grau de mestre em Economia do Setor Público, área de concentração em **Economia do Setor Público**, **Pedro de Souza Fleury**. Além dos examinadores e do examinando, compareceram também, representantes do Corpo Docente e do Corpo Discente. Iniciando a sessão, o professor **Hilton Martins de Brito Ramalho (orientador)**, na qualidade de presidente da Banca Examinadora, comunicou aos presentes qual o fim da reunião e os procedimentos de encaminhamento da mesma. A seguir, concedeu a palavra ao candidato, para que fizesse oralmente a exposição do trabalho, apresentado sob o título: **“Avaliação de risco de irregularidades em processos de aposentadoria no âmbito do Tribunal de Contas do Estado da Paraíba: uma aplicação baseada em aprendizagem de máquina supervisionada”**. Concluída a exposição, o senhor presidente solicitou que fosse feita a arguição por cada um dos examinadores. Após isso, foi concedida a palavra ao candidato para que respondesse e esclarecesse as questões levantadas. Terminadas as arguições, a Banca Examinadora passou a proceder à avaliação e ao julgamento do candidato. Em seguida, o senhor presidente comunicou aos presentes que a Banca Examinadora, por unanimidade, **aprovou** a dissertação apresentada e defendida com o conceito **APROVADO**, concedendo assim, o grau de **Mestre em Economia do Setor Público** ao mestrando **Pedro de Souza Fleury**. E, para constar, eu, **Hilton Martins de Brito Ramalho**, lavrei a presente ata, que assino junto com os membros da Banca Examinadora. João Pessoa, 27 de novembro de 2024.

Prof. Dr. Hilton Martins de Brito Ramalho 
Orientador

Documento assinado digitalmente
HILTON MARTINS DE BRITO RAMALHO
Data: 27/11/2024 11:35:00-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Aléssio Tony Cavalcanti de Almeida 
Examinador Interno

Documento assinado digitalmente
ALESSIO TONY CAVALCANTI DE ALMEIDA
Data: 28/11/2024 10:50:19-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. José Jorge Dias Lima Júnior 
Examinador Externo

Documento assinado digitalmente
JOSE JORGE LIMA DIAS JUNIOR
Data: 28/11/2024 10:44:41-0300
Verifique em <https://validar.iti.gov.br>

*Aos meus pais Lúcia Maria Moura de Souza e Marcus Drummond Fleury da Fonseca,
por estarem comigo em todos os momentos.*

Agradecimentos

Primeiramente, agradeço a Deus, pela força, sabedoria e perseverança que Ele me concedeu ao longo desta jornada.

Agradeço ao meu orientador, Prof. Dr. Hilton Martins de Brito Ramalho, cuja orientação, paciência e conhecimento foram fundamentais ao longo deste processo.

À minha família, em especial em memória de minha mãe, Lúcia Maria Moura de Souza, cuja confiança e amor inabaláveis foram a base sobre a qual construí minha trajetória acadêmica, dedico esta conquista. Ao meu pai, Marcus Drummond Fleury da Fonseca, que sempre acreditou em mim e me apoiou incondicionalmente em todas as etapas da minha vida, meu eterno agradecimento. Também agradeço aos meus irmãos Daniel Fleury e Mariana Fleury pelo incentivo constante.

À minha esposa, Bárbara Ribeiro de Belmont Fonseca, minha eterna gratidão por seu amor, paciência e apoio incondicional. Você esteve ao meu lado em cada momento desta jornada, oferecendo palavras de encorajamento e sendo uma fonte constante de inspiração e força. Seu apoio foi essencial para que eu pudesse superar os desafios e alcançar este objetivo.

Ao grande amigo e Auditor do TCE/PB, Rafael Espíndola Alexandrino, pela inestimável parceria e pelas inúmeras conversas que foram essenciais na escolha do tema e na definição dos parâmetros deste estudo. Sua generosidade em compartilhar conhecimentos e sua perspicácia foram fundamentais para o desenvolvimento deste trabalho.

Aos meus amigos e colegas de curso, pela companhia nesses intensos anos de aprendizagem e dos desafios que compartilhamos.

Agradeço ainda ao Tribunal de Contas do Estado da Paraíba (TCE/PB), não apenas por ter viabilizado a realização deste mestrado em instituição tão renomada, mas também pelo fornecimento de dados e informações essenciais para a realização da presente pesquisa.

Por fim, gostaria também de expressar minha gratidão aos professores e funcionários da UFPB, sob a tutela da Professora Dra. Maria da Conceição Sampaio de Sousa.

*“Não to mandei eu? Esforça-te, e tem bom ânimo; não te atemorizes, nem te espantes;
porque o Senhor teu Deus está contigo, por onde quer que andares.
(Josué 1:9)*

Resumo

Com o acelerado aumento do peso no orçamento público que as aposentadorias têm assumido, a cada ano, tem se tornado cada vez mais estratégica a eficiência na gestão dos recursos despendidos pelos institutos de previdência neste tipo de benefício previdenciário. O presente estudo aborda a problemática da crescente demanda por auditorias nos processos de concessão de aposentadoria e a limitação de recursos humanos disponíveis para tal fim. Utilizando dados históricos e estruturados referentes aos processos de aposentadoria enviados ao TCE/PB pelos institutos de previdência, a pesquisa visa avaliar se há modelo capaz de identificar, de forma prévia e automatizada, o risco intrínseco de irregularidades em processos de aposentadoria, de modo a permitir a priorização da análise a ser feita pelo Tribunal de Contas da Paraíba. Para tanto, são treinados onze diferentes modelos de classificação binária, com segmentação da base de testes de acordo com o grau de consenso entre os modelos. Para cada processo, além da classificação individual derivada de cada modelo, também foi considerada a escolha coletiva, derivada da classificação majoritária realizada pelos 11 modelos. A avaliação de desempenho dos modelos e da escolha coletiva foi guiada por meio de métrica de benefício-custo, a qual busca ponderar, de uma só vez, os ganhos de eficiência decorrente dos acertos de classificação dos processos sem risco (verdadeiros negativos) e as perdas decorrentes da classificação incorreta de processos que, em verdade, tem risco (falsos negativos). A avaliação derivada de tal métrica indica que, para os processos em que houve consenso de 9 ou mais modelos, a automatização apresenta performance superior à escolha puramente aleatória realizada pela Auditoria. Em contrapartida, para os processos em que 8 ou menos modelos concordaram na classificação, a escolha aleatória apresenta performance superior. Nesse contexto, verificou-se que a solução ótima é a aplicação de uma estratégia mista, com aplicação da escolha coletiva para os processos em que 9 ou mais modelos concordam e com aplicação da escolha aleatória para os processos em que 8 ou menos modelos concordam. Tal combinação proporciona um Benefício-Custo médio resultante de 2,197, índice 64,0% superior à escolha puramente aleatória atualmente realizada pelo TCE/PB.

Palavras-chaves: Aposentadoria. Irregularidades. Risco. Auditoria. Aprendizagem Supervisionada. Eficiência.

Abstract

With the accelerated increase in the share of public budgets allocated to retirees each year, the efficient management of resources spent by retirement funds on this type of social benefit has become increasingly strategic. This study addresses the growing demand for audits in retirement benefit approval processes and the limitations of available human resources to meet this demand. Using historical and structured data related to retirement processes submitted to the State Court of Auditors of Paraíba (TCE/PB) by pension institutes, the research seeks to evaluate whether a model can automatically and preliminarily identify the intrinsic risk of irregularities in retirement processes, thereby enabling the prioritization of cases for analysis by the Court. To achieve this, eleven different binary classification models are trained, with the test dataset segmented according to the degree of consensus among the models. For each process, in addition to the individual classification derived from each model, the collective choice derived from the majority classification of the 11 models was also considered. The performance evaluation of the models and the collective choice was guided by a benefit-cost metric, which seeks to simultaneously weigh the efficiency gains from correctly classifying low-risk processes (true negatives) and the losses from incorrectly classifying processes that, in reality, carry risk (false negatives). The evaluation based on this metric indicates that, for processes where 9 or more models reached consensus, automation demonstrates superior performance compared to the purely random selection performed by the Audit. Conversely, for processes where 8 or fewer models agreed on the classification, the random choice shows better performance. In this context, the optimal solution was found to be a mixed strategy, applying the collective choice for processes where 9 or more models agree and random selection for processes where 8 or fewer models agree. This combination yields an average Benefit-Cost ratio of 2.197, which is 64.0% higher than the purely random selection currently employed by the TCE/PB.

Keywords: Retirement. Irregularities. Risk. Audit. Supervised Learning. Efficiency.

Lista de tabelas

Tabela 1 – Variáveis da base de dados consolidada	21
Tabela 2 – Variáveis da base pré-treinamento	26
Tabela 3 – Conceito de matriz de confusão	31
Tabela 4 – BC da escolha coletiva conforme grau de consenso dos modelos . . .	40
Tabela 5 – BC da escolha de cada modelo para os processos em que 11 modelos concordaram	41
Tabela 6 – BC da escolha de cada modelo para os processos em que 10 modelos concordaram	41
Tabela 7 – BC da escolha de cada modelo para os processos em que 9 modelos concordaram	43
Tabela 8 – BC da escolha de cada modelo para os processos em que 8 ou menos modelos concordaram	45
Tabela 9 – BC resultante da estratégia mista entre escolha coletiva e a escolha puramente aleatória	48
Tabela 10 – BC resultante da estratégia mista entre a melhor escolha individual agregada e a escolha puramente aleatória	49
Tabela 11 – Quantidade de variáveis dummy por campo categórico	50
Tabela 12 – Grau de importância de cada atributo de treinamento	52
Tabela 13 – Palavras-chave e termos relacionados para revisão da literatura . . .	63

Lista de ilustrações

Figura 1 – Processo de concessão de aposentadoria nos institutos de previdência 11

Lista de abreviaturas e siglas

ALICE	Análise de Licitações e Editais
AED	Análise Exploratória de Dados
BC	Benefício-Custo
FN	Falso Negativo
FP	Falso Positivo
GEO/PB	Georreferenciamento de Obras da Paraíba
IDGPB	Indicadores de Desempenho do Gasto Público em Educação e Saúde na Paraíba
KNN	K-Nearest-Neighbour
RPPS	Regime Próprio de Previdência Social
SAGRES	Sistema de Acompanhamento da Gestão dos Recursos da Sociedade
SVM	Support Vector Machine
TI	Tecnologia da Informação
TRAMITA	Sistema de Tramitação de Documentos do TCE/PB
TCE	Tribunal de Contas do Estado
TCE/PB	Tribunal de Contas do Estado da Paraíba
TCM	Tribunal de Contas dos Municípios
TCU	Tribunal de Contas da União
UFPB	Universidade Federal da Paraíba
VA	Variável Aleatória

VN Verdadeiro Negativo

VP Verdadeiro Positivo

Sumário

1	INTRODUÇÃO	1
1.1	Do uso de ferramentas de TI na fiscalização de controle externo	3
1.2	Motivação	6
1.3	Objetivos	8
1.3.1	Objetivo Geral	9
1.3.2	Objetivos Específicos	9
2	O PROCESSO DE APOSENTADORIA E A ATUAÇÃO DO TCE/PB	10
3	REVISÃO DA LITERATURA	16
4	METODOLOGIA	20
4.1	Dados	21
4.2	Dos modelos de classificação e da escolha coletiva	28
4.3	Métricas gerais de avaliação de desempenho	31
4.4	Métrica de Benefício-Custo (BC)	34
5	DOS RESULTADOS	37
5.1	Do indicador BC da situação atual	37
5.2	Do indicador BC da automatização	38
5.3	Da importância dos atributos	50
6	CONSIDERAÇÕES FINAIS	54
	REFERÊNCIAS	57
	ANEXOS	61
	ANEXO A – PROTOCOLO DE REVISÃO DA LITERATURA	62

1 Introdução

Diante dos desafios inerentes à gestão da coisa pública no cenário brasileiro, a Constituição Federal de 1988 elegeu o Tribunal de Contas como órgão técnico integrante do sistema de controle externo titularizado pelo Poder Legislativo¹, responsável não só pela fiscalização de todas as etapas do processo de realização das despesas públicas², mas também por apreciar, para fins de registro (homologação), a legalidade das aposentadorias de servidores públicos efetivos concedidas pelas unidades gestoras dos regimes próprios de previdência, também conhecidos como institutos de previdência.

Tal sistema é imprescindível para a concretização do Princípio Republicano, fundamento intrínseco da organização política em um Estado democrático, que se baseia na ideia de “coisa pública”, em que o poder emana do povo e é exercido em benefício do bem comum. No contexto republicano, há uma ênfase na representação política, na participação cidadã e na responsabilização dos governantes e gestores públicos perante a coletividade, sendo esta última uma das principais funções das Cortes de Contas.

Nesse diapasão, o sistema de Controle Externo brasileiro é composto por 33 Tribunais de Contas, divididos em³:

- a) 1 Tribunal de Contas da União (TCU): Órgão federal com jurisdição sobre órgãos e entidades federais;
- b) 26 Tribunais de Contas dos Estados (TCEs): Órgãos estaduais com jurisdição sobre órgãos e entidades estaduais e municipais⁴;
- c) 3 Tribunais de Contas dos Municípios (TCMs)⁵: Órgãos estaduais com jurisdição

¹ CRFB, art. 71. Consulte em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm

² A realização de despesas públicas compreende diversas etapas: planejamento orçamentário, licitações, os contratos, empenhos, liquidações e desembolsos realizados por cada entidade componente de sua esfera de competência fiscalizatória.

³ Para mais detalhes, consulte em: <https://www.tcmgo.tc.br/site/2022/01/17-01-2022/>

⁴ Exceto se o Estado tiver também Tribunal de Contas dos Municípios, neste caso o TCE fiscaliza os órgãos e entidades estaduais e o TCM fiscaliza os órgãos e entidades municipais. Para mais detalhes, consulte: <https://enciclopediajuridica.pucsp.br/verbete/20/edicao-1/tribunais-de-contas-no-brasil>

⁵ Apenas os Estados da Bahia (BA), Goiás (GO) e Pará (PA) possuem TCMs;

sobre os municípios daquele Estado;

- d) 2 Tribunais de Contas Municipais: Órgãos municipais com jurisdição apenas sobre os órgãos e entidades daqueles municípios⁶.

Especificamente no que concerne ao Estado da Paraíba, a atuação do controle externo no monitoramento dos gastos públicos assume papel especialmente relevante, tendo em vista que se trata de umas das unidades federativas mais pobres do país⁷. Conforme dispõe o Instituto Brasileiro de Geografia e Estatística - IBGE⁸, a Paraíba ocupa apenas a penúltima colocação no ranking de PIB per capita, com valor equivalente a apenas 50% da média nacional, superando apenas o Estado do Maranhão.

Nesse sentido, destaca-se a importância das atividades realizadas pelo Tribunal de Contas do Estado da Paraíba (TCE/PB), órgão que tem sob sua jurisdição o próprio Estado, os 223 municípios, além das demais pessoas jurídicas integrantes da Administração Indireta destes entes federados, incluindo setenta e um institutos de previdência.

Tais entidades previdenciárias exercem a função de unidades gestoras dos regimes próprios de previdência do Estado e dos municípios da Paraíba, responsáveis por acumular recursos oriundos das contribuições do respectivo ente federativo e dos servidores, bem como custear os benefícios previdenciários, entre os quais o mais representativo é a aposentadoria.

A aposentadoria desempenha um papel crucial na economia do setor público, especialmente em países que já enfrentam um processo de envelhecimento populacional, como é o caso do Brasil⁹. À medida que as sociedades se tornam mais longevas, há uma transformação significativa na estrutura demográfica, com um aumento expressivo na proporção de idosos em relação à população economicamente ativa. Essa mudança demográfica gera impactos profundos nos sistemas previdenciários e no orçamento público, exigindo atenção especial dos gestores e formuladores de políticas públicas.

⁶ A criação de novas Cortes de Contas municipais foi vedada pela CF/88, mantendo-se apenas os Tribunais desse tipo já existentes à época: município do Rio de Janeiro e município de São Paulo.

⁷ A Paraíba ocupa o 23º lugar no ranking de IDH dos estados brasileiros. Para mais informações sobre o ranking, consulte em: <http://www.atlasbrasil.org.br/ranking>

⁸ Para mais informações, consulte o IBGE: https://biblioteca.ibge.gov.br/visualizacao/livros/liv102045_informativo.pdf

⁹ Veja o estado do envelhecimento populacional brasileiro, conforme o Censo de 2022: <https://www.gov.br/secom/pt-br/assuntos/noticias/2023/10/censo-2022-numero-de-idosos-na-populacao-do-pais-cresceu-57-4-em-12-anos>

A garantia de um sistema de aposentadoria sustentável é essencial não apenas para proteger o bem-estar social dos indivíduos que atingem a inatividade laboral, mas também para preservar a estabilidade econômica do setor público. No entanto, o envelhecimento populacional impõe desafios crescentes. Com um número cada vez maior de beneficiários e um número proporcionalmente menor de contribuintes, os sistemas previdenciários enfrentam pressão fiscal significativa, aumentando o peso das despesas com aposentadorias no orçamento público. Esse cenário é agravado em países onde os sistemas previdenciários são majoritariamente baseados em regimes de repartição, em que os trabalhadores ativos financiam as aposentadorias dos inativos, como é o caso, em boa parte, da previdência pública brasileira¹⁰.

Além disso, o aumento das despesas com aposentadorias pode comprometer a capacidade do setor público de alocar recursos para outras áreas essenciais, como saúde, educação e infraestrutura. Assim, a gestão eficiente desses sistemas torna-se fundamental para equilibrar a proteção social dos aposentados com a sustentabilidade fiscal de longo prazo, razão pela qual a análise da regularidade das aposentadorias concedidas pelos institutos de previdência assume especial importância para a economia do setor público paraibano.

À luz de todo o exposto, a presente pesquisa tem como enfoque, justamente, avaliar a viabilidade de automatização da avaliação de risco de irregularidades em aposentadorias concedidas pelos institutos de previdência do Estado e municípios da Paraíba aos seus servidores, atividade esta que é exercida, por direta atribuição constitucional¹¹, pelo Tribunal de Contas do Estado da Paraíba.

1.1 Do uso de ferramentas de TI na fiscalização de controle externo

A intensificação da utilização de ferramentas de Tecnologia da Informação (TI) e Comunicação alterou significativamente a forma de trabalho e de interação da sociedade. Neste sentido, desde o início da década de 2000, o Governo Federal tem avançado no desenvolvimento do Programa de Governo Eletrônico (e-Gov), por meio do qual, à

¹⁰ Entenda mais sobre os regimes previdenciários brasileiros: <https://www.jusbrasil.com.br/artigos/regimes-da-previdencia-no-brasil/830099724>

¹¹ Constituição da República Federativa do Brasil, art. 71, inciso III.

semelhança do que ocorre nas relações privadas, tem propiciado transformação relevante nas relações do Governo para com seus cidadãos, as empresas e seus próprios órgãos¹².

Parte deste processo vem ancorado em um esforço de desenvolvimento paulatino da infraestrutura de TI, de software e da integração de bases de dados e sistemas que permeiam o Governo Federal, o que, apesar das enormes lacunas e desafios ainda existentes, têm influenciado a forma que o Tribunal de Contas da União atua na fiscalização dos órgãos e entidades da Administração Pública Federal.

Para isso, a Corte de Contas Federal desenvolveu o Laboratório de Informações e Controle – LabContas, ambiente virtual que permite acesso aos servidores que atuam na fiscalização de mais de 100 bases de dados custodiadas pelo TCU, além de ferramentas de consultas, análise e cruzamento de dados. Utilizando a grande gama de informações disponíveis, observa-se um esforço em automatizar diversas verificações de irregularidades nas mais variadas áreas objeto de sua atuação, a partir de ferramentas digitais que têm gerado excelentes resultados (COSTA; BASTOS, 2020).

Por exemplo, desde 2017 encontra-se operante o sistema ALICE, acrônimo para Análise de Licitações e Editais, um robô que processa o conteúdo de todos os editais de licitações e atas de registros de preços publicados por órgãos e entidades da Administração Pública Federal. Em outras palavras, trata-se de uma ferramenta fundamental para apresentar indícios de irregularidades aos Auditores de Controle Externo em procedimentos de licitação, de maneira que possam ser corrigidos por uma atuação preventiva da Corte de Contas Federal, antes da efetiva ocorrência da licitação.

Há outros exemplos, como o Mônica, Adele, Sofia, Carina e Ágata (COSTA; BASTOS, 2020), ferramentas digitais concebidas no LabContas e que têm ajudado a racionalizar suas atividades, melhorar a eficiência na utilização da força de trabalho dos seus Auditores e possibilitar uma fiscalização orientada a dados, em sintonia com a Estratégia Digital aprovada em 16/09/2020 pelo Plenário da Corte Federal¹³.

No contexto do Estado da Paraíba, embora não se tenha o desenvolvimento alcançado a nível federal em termos de Tecnologia da Informação, já há avanços importantes, muitos deles por iniciativa do Tribunal de Contas do Estado.

¹² Consulte a linha do tempo do Programa de Governo Eletrônico Federal em: <https://www.gov.br/governodigital/pt-br/estrategia-de-governanca-digital/do-eletronico-ao-digital>

¹³ Consulte a estratégia digital em: <https://sites.tcu.gov.br/estrategia-digital/>

Em primeiro lugar, a Corte Paraibana é uma das que se destacam no cenário nacional quando o assunto é transparência, não só conferindo amplo acesso à sociedade aos dados não sigilosos que dispõe¹⁴, mas também exigindo que os jurisdicionados enviem ao tribunal dados referentes a cada etapa do planejamento e da realização de receitas e despesas¹⁵.

Nesse norte, diante da grande quantidade de dados que recebe dos jurisdicionados, aliado às informações obtidas por convênios de compartilhamento de dados pactuado com outros órgãos da Administração Pública de todas as esferas da federação, cria-se um ambiente propício para ação inovadora em diversas áreas de sua competência, o que vem ocorrendo gradativamente.

Destarte, desde 2017, estão à disposição dos Auditores de Controle Externo diversos painéis que permitem a realização de consultas e a visualização de irregularidades atinentes a contextos específicos. São dignos de nota o Painel de Medicamentos¹⁶, o de Licitantes 100% Perdedores¹⁷ e o de Acumulação de Vínculos¹⁸.

Além desses, há a ferramenta Preço de Referência¹⁹, plataforma de cotações para compras públicas desenvolvida por servidores do Tribunal em conjunto com a Universidade Federal da Paraíba e o Governo do Estado, onde são apresentados parâmetros de mercado adequados às compras públicas em variados contextos, possibilitando aos jurisdicionados a formação de mapas de preços que balizem de forma apropriada o procedimento licitatório, além de facilitar a fiscalização da adequação dos valores praticados.

No contexto do uso de inteligência artificial, deve ser ressaltado o Sistema AJUNTA²⁰, plataforma concebida pelos técnicos do TCE/PB que processa toda a base de documentos da Corte de Contas e identifica relações de parentesco de servidores, indícios de conluio em

¹⁴ Sistemas TRAMITA e SAGRES Online são franqueados à consulta em plataforma web por qualquer cidadão.

¹⁵ Entre outras informações, são obrigatoriamente informados periodicamente ao Tribunal: Projetos das leis orçamentárias, leis orçamentárias, arrecadação de receitas, avisos de licitações, editais, contratos, empenhos, liquidações, pagamentos, saldos de disponibilidades, inscrição em restos a pagar, transferências financeiras, fornecedores, folhas de pagamento.

¹⁶ Consulte o painel de medicamentos em: <https://sagres.tce.pb.gov.br/paineis-medicamentos/>

¹⁷ Consulte o painel de licitantes 100% perdedores em: <https://tce.pb.gov.br/paineis/licitantes-perdedores>

¹⁸ Consulte o painel de acumulação de vínculos em: <https://tce.pb.gov.br/paineis/acumulacao-de-vinculos-publicos>

¹⁹ Consulte a plataforma no sítio: <https://precodereferencia.tce.pb.gov.br/>

²⁰ A ferramenta é restrita ao uso pelos servidores do Tribunal em trabalhos de auditoria.

licitações, além de outros padrões e correlações úteis para priorização de ações preventivas e corretivas do Tribunal.

Por fim, outros trabalhos têm sido desenvolvidos internamente ou em parceria com a UFPB, tais como o Banco de Legislação²¹, a Turmalina²², os Indicadores de Desempenho do Gasto Público em Educação e Saúde na Paraíba (IDGPB)²³, o Sistema de Georreferenciamento de Obras (Geo-PB)²⁴, entre outros.

Diante dessas tendências, a proposta desta pesquisa é oferecer uma contribuição à Corte de Contas Paraibana que lhe possibilite dar mais um passo em seu processo de tomada de decisão orientada a dados, pautando as atividades de fiscalização no âmbito dos processos previdenciários à avaliação de risco suportada por tecnologias de aprendizagem de máquina, buscando-se maior racionalização do uso da mão de obra disponível e remodelação do fluxo processual.

1.2 Motivação

Para se ter uma ideia da ordem de grandeza das atividades sob a fiscalização do TCE/PB, apenas no ano de 2022 foram gerados, somando-se todos os seus jurisdicionados, 224 orçamentos, 35.915 licitações, 2.332.171 empenhos, conforme dados obtidos do Sistema de Acompanhamento e Gestão dos Recursos da Sociedade - SAGRES²⁵ e do sistema de tramitação de processos e documentos do TCE/PB - TRAMITA²⁶, além de denúncias, prestações de contas, aposentadorias, pensões e inspeções especiais.

Embora haja enorme gama de possíveis atos a serem fiscalizados pelo TCE/PB, apenas uma parcela deles de fato acarreta a formalização de processos que serão efetivamente analisados pelos auditores da Corte. Isso ocorre pela amostragem realizada pela Corte de Contas Paraibana, a partir da aplicação de matriz de risco que seleciona as licitações, os contratos, e, por conseguinte, as despesas que se enquadrem em tipologias

²¹ Consulte os atos normativos em: <https://legislacao.tce.pb.gov.br/>

²² Utilize o Robô Turmalina, referência em avaliação de portais de transparência em: <https://turmalina.tcepb.tc.br/turmalina/mapleaf>

²³ Consulte em: <https://idgpb.tce.pb.gov.br/>

²⁴ Saiba mais sobre a ferramenta em: <https://tce.pb.gov.br/noticias/geo-pb-ferramenta-de-controle-de-obras-e-servicos-de-engenharia>

²⁵ O sistema SAGRES pode ser consultado em: <https://sagresonline.tce.pb.gov.br/#>

²⁶ O sistema TRAMITA pode ser consultado em: <https://tramita.tce.pb.gov.br/tramita/pages/main.jsf>

previamente determinadas pelo setor de inteligência²⁷ como indicadoras de maiores riscos de lesão ao patrimônio público. Os demais tipos objetos de fiscalização, no entanto, são analisados exaustivamente pelo Tribunal, merecendo realce os processos de registro de aposentadoria, pela expressiva demanda exigida do corpo técnico de auditores.

Para que se possa compreender a relevância deste tema para a organização administrativa do TCE/PB, observa-se que, desde o exercício de 2019, foram criados 30.676 processos passíveis de análise pela auditoria e de julgamento pelo Tribunal, dos quais 12.590 são relativos a atos de concessão de aposentadoria²⁸. Em outras palavras, 41% da demanda total de processos que foram criados nos últimos cinco anos se refere a um tipo específico, ao passo que apenas 8,5% da força de trabalho (Auditores e Técnicos) está disponível para a realização deste tipo de fiscalização, denotando descompasso entre a demanda de processos e a oferta de trabalho²⁹.

Por oportuno, destaca-se que os processos de aposentadoria, ao serem formalizados pelos institutos de previdência no sistema do TCE/PB, são compostos de uma parcela de dados não estruturados e outra de dados estruturados, esta última até então não explorada por técnicas de aprendizagem de máquina (*machine learning*).

Nesse contexto, diante da elevada representatividade de tal tema para a alocação de recursos pelo Tribunal e da existência de significativa base de dados à disposição, a presente pesquisa procura avaliar a viabilidade de, por meio de técnicas de aprendizagem supervisionada, realizar, antes de qualquer análise técnica pela Auditoria, a classificação de risco de cada novo processo de aposentadoria, a partir do uso de informações estruturadas de milhares de processos já julgados pela Corte constantes de seu banco de dados, cujo rótulo para treinamento será estabelecido em função da quantidade de inspeções ou relatórios que cada processo demandou da Auditoria.

Assim, ao se considerar as limitações do quadro de pessoal alocado nas atividades de controle externo, algoritmos baseados em aprendizagem de máquina surgem, portanto, como elementos de papel estratégico no cumprimento da missão constitucional dos tribunais de contas, auxiliando não só na coleta, no armazenamento de dados e no provimento

²⁷ Gerência da Informação - GI.

²⁸ Dados obtidos em consulta ao sistema TRAMITA em março de 2024.

²⁹ Tal desequilíbrio é um pouco mitigado pelos Auditores que estão em teletrabalho, tendo em vista que as metas extraordinárias para fruição desse regime se traduzem em realizar 15 processos de aposentadoria por mês.

de ferramentas de otimização e racionalização da força de trabalho, mas, sobretudo, habilitando a tomada de decisões orientada a dados, com a priorização dos esforços da Organização em função de fatores como risco, materialidade e relevância.

Tanto a expressividade dos processos que constituem o escopo deste trabalho como a realidade estratégica da Instituição demonstram o potencial contributivo da pesquisa, uma vez que a motivação específica do projeto é avaliar se há um modelo (ou conjunto de modelos) capaz de classificar, a partir de informações constante dos dados estruturados, os processos de aposentadoria recém formalizados pelo Tribunal utilizando aprendizagem supervisionada, com treinamento a partir de base de dados relevante para a análise, isto é, de processos julgados pela Corte Paraibana.

Nesse contexto, a pesquisa terá como foco avaliar se é possível extrair, antes de qualquer instrução ou manifestação da Auditoria, o nível de risco de um processo utilizando-se exclusivamente dos dados estruturados enviados pelos institutos de previdência. Nesse contexto, se estabelece a primeira questão de pesquisa (QP1):

QP1: *Os dados estruturados relativos aos processos de aposentadoria enviados ao TCE/PB pelos institutos de previdência são suficientes para uma classificação, a priori, confiável de risco?*

Em sendo positiva a resposta à QP1, surge a segunda questão de pesquisa (QP2):

QP2: *Qual modelo ou agrupamento de modelos de aprendizagem supervisionada mais adequado ao objetivo da pesquisa?*

1.3 Objetivos

Dentro do contexto de transição entre o paradigma das análises formais para o que envolve as organizações orientadas a dados (*data-driven*), cuja definição do risco deriva das diversas técnicas de extração de conhecimento de bases de dados - KDD (*knowledge-discovery in databases*), são definidos alguns objetivos para a presente pesquisa, a seguir descritos.

1.3.1 Objetivo Geral

O objetivo central desta pesquisa é analisar a viabilidade da definição de um modelo (ou agrupamento de modelos) de avaliação de risco de irregularidades em processos de aposentadoria baseado em aprendizagem de máquina supervisionada.

1.3.2 Objetivos Específicos

No que concerne aos objetivos específicos, destacam-se os seguintes:

1. Identificar características e avaliar a qualidade dos dados enviados pelos institutos de previdência ao TCE/PB;
2. Identificar atributos importantes para a automatização;
3. Realizar transformações necessárias nos dados;
4. Definir a métrica de desempenho adequada;
5. Escolher os modelos a serem aplicados;
6. Definir o parâmetro de referência da métrica de desempenho na situação atual;
7. Avaliar os resultados em face dos parâmetros da situação atual;
8. Identificar as contribuições da pesquisa.

2 O processo de aposentadoria e a atuação do TCE/PB

Com intuito de facilitar o entendimento do tema de pesquisa, é fundamental apresentar um breve resumo das etapas trilhadas por um servidor público até se aposentar, bem como especificar o papel do TCE/PB em relação às aposentadorias concedidas.

O vínculo funcional do servidor público efetivo se inicia com a investidura em cargo público após aprovação em concurso público de provas ou de provas e títulos. A partir da posse, inicia-se o vínculo da pessoa com o ente federativo e, igualmente, o servidor passa a também estar vinculado ao respectivo instituto de previdência gestor do Regime Próprio de Previdência Social (RPPS).

Ao longo de sua trajetória como ocupante de cargo público efetivo, o servidor contribui mensalmente para o instituto de previdência, até que adquira o direito de se aposentar pelo atingimento de alguma das regras constitucionais aplicáveis.

No momento em que o servidor solicita sua aposentadoria, o respectivo instituto de previdência avalia o atendimento de todos os requisitos constitucionais e legais e, ao concordar com o pleito, publica uma portaria de concessão do benefício, transferindo o servidor para a inatividade.

Ocorre que o ato de concessão de aposentadoria é considerado um ato administrativo complexo, que somente é finalizado após o julgamento de sua legalidade pelo Tribunal de Contas competente, razão pela qual o instituto de previdência tem de formalizar perante o TCE/PB processo individual referente a cada aposentadoria que tenha concedido.

Vejamos o fluxograma deste processo:

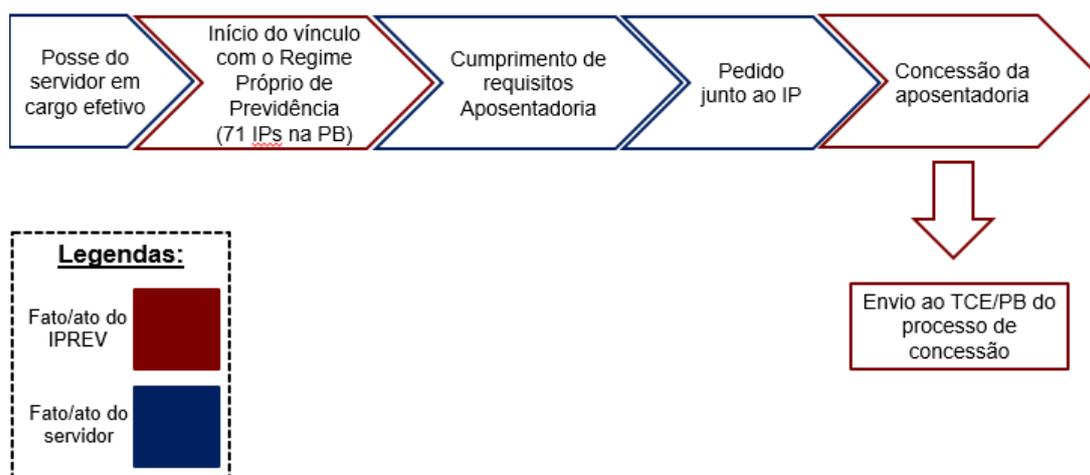


Figura 1 – Processo de concessão de aposentadoria nos institutos de previdência

Essa formalização é feita a partir do envio de diversos documentos comprobatórios em formato *portable document format* - pdf, nos quais constam as seguintes informações:

1. Requerimento do(a) interessado(a) ou de seu representante legal com pedido de concessão do benefício;
2. Documentos pessoais do(a) servidor(a): CPF, documento de identificação com foto e documento que identifique o estado civil atual do(a) mesmo(a);
3. Ato de provimento do(a) servidor(a) civil ou militar para o cargo efetivo em que se der a aposentadoria, reserva, reforma ou o fundamento para a pensão;
4. Ficha funcional do(a) servidor(a) civil ou militar com os respectivos assentamentos;
5. Demonstrativo de Tempo de Contribuição no modelo adotado no sistema do Ministério da Previdência;
6. Fichas financeiras a partir de julho de 1994 ou do início da contribuição, se posterior àquela competência;
7. Parecer jurídico;
8. Memória de cálculo dos proventos com base na média ou na última remuneração do(a) servidor(a), conforme o caso;

9. Ato concessório contendo o nome, matrícula, cargo (para servidores civis) e nome dos pensionistas (em caso de pensão), graduação ou posto (para militares), lotação, modalidade de aposentadoria, reserva, reforma ou pensão, fundamento jurídico do benefício, data e assinatura da autoridade competente;
10. Comprovante de publicação do ato concessório em órgão oficial de imprensa;
11. Comprovação da implementação dos cálculos nos proventos do(a) servidor(a) aposentado(a) / reservista / reformado(a) / pensionista.

Além das informações não estruturadas, é enviada quantidade significativa de dados estruturados, contendo informações relevantes do processo daquele servidor, a saber:

1. Tipo de aposentadoria (geral, compulsória ou por invalidez);
2. Identificação do instituto de previdência;
3. Exercício de concessão do benefício;
4. Identificação da carreira e do cargo que o servidor atuou por último;
5. Órgão de lotação do servidor em seu último vínculo;
6. Data de ingresso na carreira;
7. Data da publicação da concessão da aposentadoria;
8. Identificação sobre se é um processo de aposentadoria especial (como por exemplo fazem jus os professores);
9. Descrição da regra pela qual a pessoa está se aposentando;
10. Número de dias que o servidor laborou em cada vínculo componente de seu tempo de contribuição;
11. Datas inicial e final em cada vínculo;
12. Identificação se cada vínculo era privado ou público;
13. Identificação do nome da empresa ou órgão correspondente a cada vínculo;

14. Quantidade de dias fictos aplicáveis a cada vínculo, se couber;
15. Quantidade de dias de magistério aplicáveis a cada vínculo, se couber;
16. Nome de cada parcela que o aposentado está recebendo a título de proventos pelo instituto de previdência;
17. Nome de cada parcela componente da última remuneração do cargo efetivo que ocupava antes de se aposentar;
18. Valor médio das remunerações do servidor ao longo de sua vida funcional, quando se tratar de servidor que se aposentou por regra que leva em conta a média;
19. Valor da remuneração do cargo efetivo antes de se aposentar.

Tendo em vista a baixa qualidade dos arquivos em formato “pdf” que são enviados pelos institutos de previdência, a maior exigência para processamento deste tipo de arquivo e a dificuldade inerente à interpretação automatizada de imagens, o presente projeto tem como foco avaliar se é possível extrair, antes de qualquer instrução ou manifestação da Auditoria, o nível de risco/complexidade de um processo, utilizando-se exclusivamente dos dados estruturados preenchidos pelos institutos de previdência.

Com intuito de responder a QP1¹, é necessário estabelecer qual é o critério de risco existente no histórico de processos existente, para fins de definição do rótulo para treinamento e teste dos modelos.

Para isso, considera-se apropriado utilizar a quantidade de relatórios de Auditoria que foram necessários na instrução de cada processo inerente à base de treinamento da ferramenta. Dito de outra forma, se um processo foi julgado após apenas um relatório de auditoria, isso significa que não há irregularidades que tenham exigido explicações ou ajustes complementares pelo instituto de previdência. Ao revés, se um processo teve mais de um relatório de auditoria, isso significa que há maior risco no processo subjacente, sendo, então, o critério de risco diretamente proporcional à quantidade de relatórios técnicos emitidos antes do julgamento definitivo. A partir dessa avaliação, é possível a atribuição de um rótulo binário a cada processo que constitui a base de dados, o qual

¹ **QP1:** *Os dados estruturados relativos aos processos de aposentadoria enviados ao TCE/PB pelos institutos de previdência são suficientes para uma classificação, a priori, confiável de risco?*

guiará a etapa de treinamento dos modelos e de avaliação de performance classificatória da base de testes.

A partir da avaliação de diversos modelos apontados nesta proposta, será avaliado o desempenho, a partir de critérios mais à frente descritos, possibilitando-se responder a QP2².

Por oportuno, é importante destacar que em 2019 foi promulgada e publicada uma nova reforma da previdência por meio da Emenda Constitucional nº 103, que trouxe uma série de modificações ao sistema previdenciário brasileiro, com novas regras e parâmetros a serem adotados tanto pelo Regime Geral de Previdência Social (RGPS) quanto pelos Regimes Próprios de Previdência Social (RPPS).

Por conseguinte, o Tribunal de Contas do Estado da Paraíba precisou aguardar a publicação de normativos específicos por parte dos municípios que compõem o Estado, sendo que apenas recentemente esse processo de transição foi concluído, o que ofereceu a oportunidade de a Instituição alterar a parametrização de seus sistemas a fim de compatibilizá-los às novas possibilidades de dados.

Nesse cenário, entre 2020 e meados de 2023, os processos de aposentadoria em que se aplicam as novas regras foram enviados ao TCE/PB sem a adaptação necessária à nova realidade, fazendo com que esse histórico não seja suficientemente relevante, dado que existem muitas informações não validadas (processos ainda não auditados) e em volume insuficiente para o tipo de aplicação pretendida neste trabalho, o que justificou o corte na base de dados, a qual apenas inclui aposentadorias concedidas sob as regras anteriores à Emenda Constitucional 103/2019.

Em que pese essa redução de escopo, as regras anteriores continuam fundamentando concessões de aposentadorias de servidores que já haviam adquirido o direito antes de novembro de 2019, embora o número de novos processos na regra anterior apresente uma redução gradativa ao longo do tempo.

Portanto, por restrição de dados confiáveis para treinamento dos modelos, o presente projeto se limita à avaliação de risco dos atos de concessão de aposentadoria nos quais se apliquem as regras anteriores à reforma da previdência de 2019, o que compõe um

² **QP2:** *Há um modelo ou agrupamento de modelos de aprendizagem supervisionada mais adequado ao objetivo da pesquisa?*

histórico de mais de 17.000 processos totalmente instruídos e julgados.

Não obstante, essa limitação não retira a validade da pesquisa, que servirá de prenúncio para futura ampliação do estudo que contemple as novas regras, em momento que houver histórico confiável de processos para treinamento dos modelos de aprendizagem supervisionada.

3 Revisão da Literatura

Conforme explicitado no protocolo de revisão da literatura (Anexo A), foram identificados 17 artigos científicos que utilizavam a aprendizagem de máquina como apoio à detecção de anomalias ou fraudes¹ em diversas áreas.

Um conjunto amplo desses estudos se concentrou na detecção de fraudes no sistema financeiro, com aplicações em mercados industriais, de cartões de crédito e de pagamento, em aplicativos de *internet banking*, bem como na detecção de lavagem de dinheiro, inclusive no âmbito das criptomoedas (RANDHAWA et al., 2018; DOMASHOVA; KRIPAK, 2021; LOKANAN; SHARMA, 2022; WEN et al., 2022; STEFÁNSSON et al., 2022; LABANCA et al., 2022; LOKANAN, 2023; FAKIHA, 2023; HAMZA et al., 2023; ASOMURA; IJIMA; MORI, 2023).

No contexto do sistema de saúde dos Estados Unidos da América – *Medicare*, destaca-se o estudo de Herland, Bauder e Khoshgoftaar (2019). Seu objetivo foi identificar transações fraudulentas provenientes das declarações feitas por clínicas/profissionais de saúde que integram a respectiva rede. Na pesquisa apresentada, demonstram-se os efeitos do grave desequilíbrio de classes e da raridade utilizando um método de avaliação de treinamento e teste (Train_Test) por meio de um conjunto de retenção, fornecendo recomendações com base nos resultados do aprendizado de máquina supervisionado. Além disso, repetem-se os mesmos experimentos usando Validação Cruzada, determinando que esta é uma alternativa viável para a detecção de fraudes no sistema Medicare. No caso de aprendizado de máquina com conjuntos de dados de grave desequilíbrio de classes, constatou-se que, conforme esperado, o desempenho na detecção de fraudes diminui à medida que as instâncias fraudulentas se tornam mais raras. Aplica-se a Subamostragem Aleatória (Random Undersampling) tanto ao Train_Test quanto à Validação Cruzada, em todos os conjuntos de dados originais e gerados, com o objetivo de avaliar possíveis

¹ Como se sabe, fraudes se materializam por qualquer intencional ardiloso, feito com o intuito de ludibriar ou lesar outrem. Todavia, é relevante destacar que a definição de irregularidades utilizadas para o presente estudo não se resume às fraudes, mas também compreende erros que, embora não intencionais, possibilitem a concessão de benefícios de aposentadoria em desacordo ao que está prescrito na legislação aplicável.

melhorias na detecção de fraudes, reduzindo os efeitos adversos do desequilíbrio de classes e da raridade. No geral, os resultados indicam que o método Train_Test supera significativamente a Validação Cruzada.

Por sua vez, Henrique, Sobreiro e Kimura (2020) empregaram técnicas de aprendizagem de máquina para verificar riscos de descumprimento de contratos firmados na seara da administração pública, dado que a interrupção de serviços essenciais pode levar a grandes problemas sociais. Como metodologia, os autores classificaram os contratos administrativos em função do risco associado, valendo-se, além do uso de ferramental estatístico, de uma lista de empresas suspensas de contratar com a administração pública, como indicador complementar do risco, sendo utilizadas as técnicas de regressão logística, *K-Nearest Neighbors - KNN*, *Random Forest*, *Support Vector Machine (SVM)*, bem como a análise de discriminante. O KNN foi o algoritmo que apresentou a melhor performance.

No âmbito de crimes contra o sistema financeiro, o estudo de Nicholls, Kuppa e Le-Khac (2021) reuniu diversas aplicações de aprendizagem profunda e de máquina no contexto das diversas fraudes financeiras existentes, categorizando-as e indicando-se os sistemas e algoritmos utilizados, assim como os atores envolvidos. Os modelos apresentados nesta revisão reforçam a necessidade de utilizar técnicas de detecção baseadas em Anomalias em Gráficos/Grupos para combater crimes financeiros. Os autores aceitam o desafio especialmente em relação à obtenção de conjuntos de dados rotulados e à expertise necessária para rotular verdades fundamentais quando estas não estão previamente disponíveis. Com o avanço das criptomoedas e seu crescente enraizamento no setor financeiro, não é surpreendente que pesquisas sobre combate à lavagem de dinheiro no contexto das criptomoedas tenham sido iniciadas. Os autores suspeitam que uma análise mais detalhada das criptomoedas e sua integração no domínio público, por parte dos órgãos fiscais e das autoridades de segurança dos diversos países, resultará em um aumento das pesquisas, especialmente na área de Detecção de Anomalias em Gráficos/Grupos.

Já nos estudos de Elsadig et al. (2022), por sua vez, a aprendizagem de máquina é exclusivamente utilizada na detecção do crime cibernético conhecido como *phishing*. Os experimentos mostraram que o método proposto alcançou uma precisão de 96,66% nos resultados, e então os resultados obtidos foram comparados com outros trabalhos da revisão de literatura. Os resultados indicaram que o método proposto foi eficiente e válido na detecção de URLs de sites de *phishing*.

Com um direcionamento voltado às aplicações de aprendizagem de máquina no setor público, Lyra et al. (2022) apresentaram uma revisão sistemática da literatura que, após avaliar e resumir diversas descobertas relacionadas a técnicas de detecção de crimes, destacou a aprendizagem de máquina como um das abordagens contemporâneas mais promissoras para lidar com fraudes, corrupção e conluio em processos de contratação pública, apresentando melhores resultados aquelas que usam dados rotulados para o processo de treinamento. A equipe do estudo avaliou e resumiu descobertas relacionadas a técnicas de detecção de crimes, baseadas principalmente em aprendizado de máquina e ciência de redes, além de estudos que utilizam índices de risco de fraude. Algumas metodologias apresentaram resultados promissores na identificação de crimes, especialmente aquelas que utilizam dados rotulados e técnicas de aprendizado de máquina. No entanto, devido à frequente indisponibilidade de dados previamente rotulados sobre casos passados, a análise por meio de ferramentas de ciência de redes tornou-se mais evidente e relevante em pesquisas exploratórias.

Já Rodríguez et al. (2022) utilizaram técnicas de aprendizagem de máquina para detectar conluio em licitações, mas a aplicação foi mais ampla, dado que comportou dados colusivos do Brasil, Estados Unidos, Suíça e Japão, evidenciando que esse tipo de fraude é um problema compartilhado por diversas nações. Destacaram que o *machine learning* ainda é incipiente nesse contexto, apesar de possuir um potencial promissor. Apresentaram como maior desvantagem dessas técnicas a quantidade de dados históricos necessários para o treinamento dos modelos, mas, por outro lado, salientaram a grande flexibilidade delas, visto que elas foram capazes de identificar com boas taxas de acerto procedimentos colusivos mesmo em situações de restrição das informações disponíveis.

Por fim, e de forma similar ao artigo anterior, Brandão et al. (2024) propuseram um *pipeline* semiautomático para a detecção de fraudes em licitações públicas. A atuação do algoritmo, em síntese, foi implementada a partir de dois módulos: um metaclassificador heurístico aplicado sobre editais de licitações e um módulo de qualidade dos dados. Ambos os módulos apresentaram resultados importantes após a prova de conceito, reforçando a relevância do método e indicando a possibilidade de automatização da fiscalização em processos licitatórios. Além disso, o método permitiu a avaliação de superfaturamento em licitações, bem como fomentou o estabelecimento de trilhas de auditoria, as quais representam possibilidades de atuação mais aprofundada e otimizada por parte de especia-

listas. Um outro ponto de relevância é que o método, ao categorizar informações públicas oriundas de arquivos com extensão .pdf, permite a construção de um banco de dados cujas informações servem de conjunto de treinamento para os diversos algoritmos de *machine learning*.

Portanto, através da revisão da literatura implementada com os parâmetros evidenciados no Anexo A, verificaram-se diversas publicações que destacaram a relevância da aplicação das técnicas de aprendizagem de máquina no contexto da detecção de anomalias e fraudes no âmbito da Administração Pública, bem como no apoio à decisão do gestor público.

Embora o uso dessas técnicas sofra limitações em razão da indisponibilidade de dados, a pesquisa revelou que a implementação delas é promissora, pois pode fomentar a otimização dos recursos envolvidos, inclusive humanos, dado o potencial de automatização de diversas tarefas e a possibilidade de extração de amostras para fiscalização dotadas de alto nível de materialidade.

Nesse contexto, considerando a relevância dos procedimentos que regem a concessão de benefícios previdenciários e que contemplam a observância de leis, o adimplemento de complexos parâmetros, bem como a sustentabilidade financeira dos regimes de previdência, vislumbra-se a relevância da contribuição oferecida, dado que ela se ampara em técnicas promissoras e amplamente utilizadas, conforme verificado na revisão da literatura apresentada.

Ademais, tendo em vista que, da revisão realizada, não foram identificadas aplicações das técnicas de *machine learning* especificamente sobre dados de natureza previdenciária, destaca-se o caráter inovador da contribuição oferecida, a qual, conforme será detalhadamente tratada em tópicos posteriores, tem potencial de aplicação no escopo da fiscalização realizada pelos Tribunais de Contas brasileiros, tanto no contexto da detecção de irregularidades associadas à concessão de benefícios previdenciários, como no âmbito da maximização das performance da fiscalização realizada sobre a matéria.

4 Metodologia

A pesquisa foi iniciada com uma revisão da literatura em que não foram identificados trabalhos científicos relacionados à aplicação de técnicas de *machine learning* na detecção de risco de irregularidades em processos de concessão de aposentadoria no âmbito do setor público. Nesse contexto, ressalta-se o caráter inovador da pesquisa, cujo objeto tenciona avaliar o potencial deste tipo de tecnologia na melhoria da eficiência da análise previdenciária realizada pelo Tribunal de Contas do Estado da Paraíba.

Posteriormente, passou-se à coleta e a estruturação de uma base de dados, com a criação de uma tabela consolidada, com todas as informações pertinentes ao projeto de pesquisa.

Em seguida, foi realizada uma ampla Análise Exploratória de Dados (AED), de forma a ser possível a melhor compreensão do domínio dos dados e inspeção dos atributos explicativos por meio dos quais são treinados os modelos. Nessa etapa foram feitas diversas transformações preparatórias das variáveis adequadas ao treinamento dos modelos, bem como a redução da dimensionalidade do sistema, através da exclusão de atributos com potencial preditivo irrelevante, a exemplo daqueles que estão fortemente correlacionados a algum outro atributo que será considerado.

Findada a etapa de preparação dos dados, realizou-se o treinamento dos modelos contemplados no estudo. Uma vez estimados os modelos, foi realizada a avaliação de desempenho individual e da classificação coletiva, obtida a partir da votação majoritária dos modelos para cada processo da base de teste, utilizando-se métrica de desempenho adequada aos objetivos do TCE/PB.

Para maiores informações técnicas, acesse o repositório do github, onde estão disponíveis todas as referências bibliográficas, os códigos utilizados, bem como htmls explicatórios sobre as etapas citadas nesta dissertação: <<https://github.com/psfleury/UFPB-MESP>>.

4.1 Dados

O conjunto de dados utilizado na pesquisa foi obtido junto à Diretoria de Tecnologia (DITEC) do Tribunal de Contas do Estado da Paraíba, compondo-se das informações relacionadas a todos os processos de aposentadoria com regras anteriores à Emenda Constitucional nº 103/2019 remetidos pelos institutos de previdência ao TCE/PB a partir do exercício de 2016, momento no qual passaram a ser informados os dados estruturados utilizados nesta pesquisa¹.

Os dados foram extraídos do sistema TRAMITA do TCE/PB em scripts “.sql”, contendo os comandos de estruturação das tabelas em banco de dados PostgreSQL e a inserção de todos os registros.

A estruturação do banco de dados, a partir dos scripts disponibilizados pelo Tribunal, a formação de uma base consolidada, a partir de ajustes e cruzamentos nas tabelas do referido banco, foram feitas a partir da linguagem R², utilizando-se o software RStudio. Essas etapas encontram-se minuciosamente detalhadas em página *web* publicada no referido repositório do Github³.

Em síntese, após seguir tais passos, obteve-se a seguinte tabela consolidada preliminar (Tabela 1), contendo 2 variáveis de identificação⁴, 34 variáveis explicativas e 1 variável de resultado:

Tabela 1 – Variáveis da base de dados consolidada

Campos	Tipo de Dados	Tabela de Origem	Tipo de Variável
id	integer	prev	Não se aplica
tipo	text	prev	Explicativa
dataaudomedico	date	prev	Explicativa

¹ Portanto, os dados compreendem benefícios previdenciários enviados ao TCE/PB pelos institutos de previdência entre 2016 e 2023, todos com regras anteriores à Reforma da Previdência de 2019

² Consulte mais sobre a linguagem no site: <https://cran.r-project.org/>

³ Consulte todas as etapas da consolidação da base e análise exploratória dos dados em: https://psfleury.github.io/UFPB-MESP/explicando_os_codigos/Projeto_Final_MESP_Parte_1.html

⁴ Tais variáveis não possuem potencial preditivo, sendo somente mantidas para permitir a identificação de qual processo no sistema está sendo analisado.

Tabela 1 – Variáveis da base de dados consolidada (*continuando*)

Campos	Tipo de Dados	Tabela de Origem	Tipo de Variável
revisao	integer	prev	Explicativa
datainicioincapacidade	date	prev	Explicativa
protocolo	integer	prev_processo	Não se aplica
jurisdicionado	text	prev_processo	Explicativa
exercicio	integer	prev_processo	Explicativa
orgao	text	prev_vinculo	Explicativa
carreira	text	prev_vinculo	Explicativa
cargo	text	prev_vinculo	Explicativa
lotacao	text	prev_vinculo	Explicativa
dataingressocarreira	date	prev_vinculo	Explicativa
datapossecargo	date	prev_vinculo	Explicativa
datapublicacaoconcessaobeneficio	date	prev_vinculo	Explicativa
aposentadoriaespecial	text	prev_vinculo	Explicativa
meiopublicacaoconcessaobeneficio	text	prev_vinculo	Explicativa
datanomeacao	date	prev_vinculo	Explicativa
legislacao	text	prev_fundamento	Explicativa
descricao	text	prev_fundamento	Explicativa
tipoato	text	prev_fundamento	Explicativa
qtd_relatorios_auditoria	integer	prev_arquivos	Explicativa
soma_dias_vinculo	integer	prev_periodo	Explicativa
qtd_vinculos	integer	prev_periodo	Explicativa
qtd_vinculos_pub	integer	prev_periodo	Explicativa
qtd_vinculos_priv	integer	prev_periodo	Explicativa
desc_empresas	text	prev_periodo	Explicativa
soma_dias_ficto	integer	prev_periodo	Explicativa
soma_dias_magisterio	integer	prev_periodo	Explicativa
qtd_vinculos_mesmo_cargo	integer	prev_periodo	Explicativa

Tabela 1 – Variáveis da base de dados consolidada (*continuando*)

Campos	Tipo de Dados	Tabela de Origem	Tipo de Variável
qtd_vinculos_mesma_carreira	integer	prev_periodo	Explicativa
nome_Proventos	text	prev_proventos	Explicativa
nome_Remuneração Cargo Efetivo	text	prev_proventos	Explicativa
valor_Média	float	prev_proventos	Explicativa
valor_Proventos	float	prev_proventos	Explicativa
valor_Remuneração Cargo Efetivo	float	prev_proventos	Explicativa
rotulo	integer	prev_arquivos	de Resultado

^a Fonte: Elaboração própria.

A partir da base consolidada, foi realizada uma profunda análise exploratória de dados que contemplou:

1. Resumo estatístico das variáveis;
2. Visualização gráfica dos atributos;
3. Avaliação de domínio;
4. Tratamento de dados ausentes;
5. Exclusão de atributos redundantes ou irrelevantes;
6. Análise de balanceamento de classes;
7. Avaliação de dimensionalidade;
8. Tratamento para evitar overfitting;
9. Engenharia de features;

10. Filtragem de dados ou remoção de outliers.

Em suma, foram observadas diversas variáveis relevantes para a aprendizagem⁵ que são preenchidas pelos institutos de previdência sem qualquer tipo de padronização, em campos textuais livres. Assim, decorreu um grande esforço de pré-processamento dos dados para que os modelos pudessem capturar padrões consistentes e úteis.

Este trabalho incluiu a normalização de textos, remoção de redundâncias, tratamento de valores ausentes e categorização de informações, passos cruciais para melhorar a qualidade dos dados e, conseqüentemente, a performance dos modelos preditivos aplicados.

No âmbito da análise exploratória de dados, em suma, foram tomadas as seguintes decisões⁶:

1. Campo 'tipo': foram excluídos os registros de aposentadoria por invalidez e aposentadorias compulsórias, tendo em vista que a representatividade de tais categorias somadas era inferior a 7% do total de registros (1286 registros).
2. Campo 'aposentadoriaespecial': como as aposentadorias de policiais, deficientes e serviços penosos perfaziam apenas 0,3% da base de dados, transformou-se tal campo em tipo binário, na qual assumiu o valor 0 para as aposentadorias comuns e 1 para aposentadoria especial, independente da especificidade;
3. Campo 'jurisdicionados': trata-se da identificação do instituto de previdência. Criou-se uma categoria residual chamada "Outros" para aqueles institutos com 300 ou menos benefícios concedidos na base de dados, com vistas a evitar superajustamento;
4. Campo 'descrição': trata-se da identificação da regra que rege a concessão da aposentadoria. Criou-se uma categoria residual chamada "Outros" para aqueles institutos com 300 ou menos benefícios concedidos na base de dados, com vistas a evitar superajustamento;
5. Campo 'orgao': neste campo, por falhas no preenchimento dos institutos, foi utilizada uma categoria residual "Outros" em 6963 registros, equivalendo a 42,3% do total de registros;

⁵ A saber: órgão, carreira, meio de publicação, cargo e proventos.

⁶ Para uma análise detalhada de tais decisões, consulte: https://psfleury.github.io/UFPB-MESP/explicando_os_codigos/Projeto_Final_MESP_Parte_1.html

6. Campo 'carreira': neste campo, por falhas no preenchimento dos institutos, foi utilizada uma categoria residual "Outros" em 2820 registros, equivalendo a 17,1% do total de registros;
7. Campo 'meiodepublicacao': neste campo, diante das muitas formas de preenchimento utilizadas pelos institutos, foram criadas 4 categorias, sendo 3 específicas (correspondendo a 88% dos registros) e uma residual (correspondendo a 12% dos registros);
8. Campo 'qtd_vinculos': o campo originalmente correspondia à contagem total de vínculos que um servidor teve ao longo de sua vida funcional até pedir a aposentadoria. Transformou-se tal campo para uma variável categórica, na qual foram identificados os intervalos relevantes;
9. Campo 'qtd_vinculos_pub': o campo originalmente correspondia à contagem total de vínculos públicos que um servidor teve ao longo de sua vida funcional até pedir a aposentadoria. Transformou-se tal campo para uma variável categórica, na qual foram identificados os intervalos relevantes;
10. Campo 'qtd_vinculos_priv': o campo originalmente correspondia à contagem total de vínculos privados que um servidor teve ao longo de sua vida funcional até pedir a aposentadoria. Transformou-se tal campo para uma variável categórica, na qual foram identificados os intervalos relevantes;
11. Campo 'cargo': diante do preenchimento falho e livre feito pelos institutos de previdência, fez-se uma ampla análise, por meio de expressões regulares, para identificar todos os cargos. Nesse contexto, foi criada uma categoria residual "Outros" para 4,8% dos registros;
12. Campo 'indicador_sm': criou-se um campo para identificar se o benefício que foi concedido é equivalente a um salário mínimo ou não. Tal informação é importante pois a Auditoria pode dar menos importância para benefícios que não possam ter seu valor reduzido, diante do valor mínimo estabelecido pela Constituição Federal;
13. Campo 'proventos': a partir do preenchimento livre realizado pelos institutos, utilizando expressões regulares, foram identificados 10 tipos distintos de categorias de

proventos pagos aos aposentados. Assim, tal campo foi desdobrado em 10 campos binários, um para cada categoria, quais sejam: Vencimento; Proventos; Vantagem Pessoal; Gratificação; ATS; Adc. de Qualificação; Insalubridade; Representação; Antecipação de Aumento; Indenizatório.

Caso o TCE/PB resolva adotar a aprendizagem supervisionada⁷ como uma forma automatizada de avaliação de risco de processos futuros de aposentadoria, é fundamental que, previamente, sejam regulamentadas as categorias que podem ser preenchidas pelos institutos de previdência para os campos relevantes, inclusive com a definição de procedimento para a inclusão de novas instâncias nas tabelas auxiliares do banco de dados, de forma a contemplar mudanças futuras.

Após a realização das etapas descritas, foram obtidos os dados de pré-treinamento, compostos por uma variável para identificação do número de protocolo de cada processo no sistema do Tribunal, 24 variáveis explicativas e uma variável de resultado, equivalente ao rótulo, listadas na Tabela 2 a seguir:

Tabela 2 – Variáveis da base pré-treinamento

Campos	Tipo de dados	Tipo de variável
nr_proc	text	Identificação
jurisdicionado	text	Explicativa
ano	integer	Explicativa
orgao	text	Explicativa
carreira	text	Explicativa
cargo	text	Explicativa

⁷ Em apertada síntese, a principal diferença entre a aprendizagem supervisionada e a não supervisionada está no tipo de dados utilizados durante o treinamento do modelo. Na aprendizagem supervisionada, o modelo é treinado com dados rotulados, ou seja, cada entrada (variáveis independentes) tem uma saída ou rótulo correspondente (variável dependente). O objetivo é que o modelo aprenda a prever essas saídas para novos dados com base nesse aprendizado. Um exemplo comum é a classificação, onde o modelo aprende a classificar novos dados em categorias previamente conhecidas. Já na aprendizagem não supervisionada, o modelo trabalha com dados não rotulados, sem saídas conhecidas, e o objetivo é descobrir padrões ou agrupamentos escondidos nos dados. Um exemplo típico é o agrupamento (clustering), onde o modelo organiza os dados em grupos com base em similaridades, sem saber previamente a que grupos eles pertencem.

Tabela 2 – Variáveis da base pré-treinamento (*continuando*)

Campos	Tipo de dados	Tipo de variável
flag_aposesp	integer	Explicativa
meio_publico	text	Explicativa
regra_apos	text	Explicativa
qtd_vinculos	text	Explicativa
qtd_vinculos_publico	text	Explicativa
qtd_vinculos_privado	text	Explicativa
qtd_vinculos_mesmo_cargo	text	Explicativa
qtd_vinculos_mesma_carreira	text	Explicativa
p_vencimento	text	Explicativa
p_atos	text	Explicativa
p_proventos	text	Explicativa
p_gratificacao	text	Explicativa
p_vantpessoal	text	Explicativa
p_insalubridade	text	Explicativa
p_adcqualificacao	text	Explicativa
p_representacao	text	Explicativa
p_antecipacao	text	Explicativa
p_indenizatorio	text	Explicativa
ind_sm	boolean	Explicativa
rotulo	boolean	de Resultado

^a Fonte: Elaboração própria.

Em seguida, passou-se à aplicação dos modelos e para a avaliação dos resultados obtidos, utilizando-se o Google Colab, a partir da linguagem Python⁸. Essas etapas encontram-se minuciosamente detalhadas em página publicada no referido repositório do Github⁹.

⁸ Consulte mais sobre a linguagem no site: <https://www.python.org/>

⁹ Consulte todas as etapas de treinamento e aplicação dos mo-

4.2 Dos modelos de classificação e da escolha coletiva

No contexto da literatura empírica pesquisada, apresentam-se, abaixo, um resumo dos modelos de classificação selecionados para aplicação nos dados objeto da pesquisa:

1. **Naive Bayes:** modelo de aprendizagem supervisionada que se baseia no Teorema de Bayes, ou seja, realiza a classificação baseada na probabilidade de pertencimento daquele registro na categoria em questão. É muito utilizado para classificações textuais em que se tenha um dataset de treino com grande quantidade de dimensões (RISH, 2001);
2. **K-Nearest-Neighbour (KNN):** modelo que armazena as ocorrências disponíveis e classifica novas ocorrências com base em alguma medida de similaridade. O valor de K determina a quantidade de itens mais próximos que serão avaliados para determinação da classe da nova ocorrência (BIAU; DEVROYE, 2015);
3. **Regressão Logística:** modelo que utiliza a função logística para compreender o comportamento de uma variável dependente binária. Como o dataset utilizado apresenta apenas rótulo binário, tal modelo é passível de utilização no presente projeto (PAMPEL, 2020);
4. **Support Vector Machine (SVM):** dado um conjunto de dados de treinamento em que cada registro é marcado como pertencente a um dos rótulos, o modelo SVM atribui novos exemplos a um rótulo ou outro, de maneira a maximizar a distância entre pontos de rótulos distintos. É um classificador linear binário não probabilístico (SCHOLKOPF; SMOLA, 2018);
5. **Decision Tree:** modelo que utiliza árvores de decisão, ou seja, mapa dos possíveis resultados de uma série de decisões relacionadas. Esse tipo de modelo leva em conta as probabilidades condicionais, os custos e benefícios de cada decisão (QUINLAN, 2024);
6. **Random Forest:** modelo que utiliza um conjunto de árvores de decisão e classifica de acordo com a categoria prevalente na maioria de tais árvores. É positivo especialmente

delos e geração de resultados em: https://psfleury.github.io/UFPB-MESP/explicando_os_codigos/Projeto_Final_MESP_Parte_2.html

para mitigar o *overfitting*¹⁰ causado pelo uso de uma única árvore de decisão (BIAU; SCORNET, 2016);

7. **Bagging ou Bootstrap Aggregating:** é uma técnica de *ensemble*¹¹ que visa melhorar a estabilidade e a acurácia de algoritmos de aprendizado de máquina. Ele gera múltiplas versões de um preditor e usa um voto para obter uma classificação agregada, reduzindo a variância e mitigando o problema de *overfitting* (BREIMAN, 1996).
8. **Multi-Layer Perceptron Classifier:** é uma implementação de redes neurais artificiais que consiste em múltiplas camadas de nós. Cada nó representa um neurônio que utiliza uma função de ativação não linear. O MLP é treinado usando o algoritmo de *backpropagation* para ajustar os pesos das conexões, tornando-se eficiente para uma ampla gama de tarefas de classificação (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).
9. **Gradient Boost:** técnica de *ensemble* que cria um modelo preditivo forte combinando vários modelos fracos de forma sequencial. Cada novo modelo corrige os erros cometidos pelos modelos anteriores, resultando em uma melhoria incremental na acurácia (FRIEDMAN, 2001);
10. **ExtraTree ou Extremely Randomized Trees:** é uma variação da árvore de decisão onde a divisão em cada nó é escolhida aleatoriamente em vez de selecionar a melhor divisão. Isso ajuda a reduzir a variância do modelo e melhora sua capacidade de generalização (GEURTS; ERNST; WEHENKEL, 2006);
11. **ADA Boost ou Adaptive Boosting:** é um algoritmo de *ensemble* que combina vários classificadores fracos para formar um classificador forte. Ele ajusta iterativamente os pesos dos classificadores com base no desempenho de cada um, focando

¹⁰ *Overfitting* em *machine learning* ocorre quando um modelo se ajusta excessivamente aos dados de treino, capturando tanto padrões reais quanto ruído ou variações irrelevantes. Como resultado, o modelo tem um desempenho excelente no conjunto de treino, mas generaliza mal para novos dados, apresentando baixo desempenho em testes ou dados não vistos. Isso ocorre porque o modelo se torna muito complexo e sensível aos dados específicos de treino, em vez de aprender os padrões gerais que podem ser aplicados a outros conjuntos de dados.

¹¹ *Ensemble*, no contexto de *machine learning*, refere-se a uma técnica que combina as previsões de múltiplos modelos para melhorar a precisão e a robustez do desempenho.

mais nos exemplos que foram classificados incorretamente em iterações anteriores (FREUND; SCHAPIRE, 1999).

A escolha dos modelos se deu, em larga medida, por dois motivos principais: a) a grande variedade de abordagens técnicas subjacentes; e b) todos estão disponíveis na biblioteca 'scikit-learn' do Python, a qual foi utilizada em sua versão 1.5.2 na execução do código.

Os modelos listados são de grande interesse para aplicação nos dados objeto da pesquisa devido à capacidade de cada um em lidar com diferentes características e desafios dos dados. Cada modelo possui suas particularidades que o tornam mais ou menos adequado para determinados cenários e tipos de dados, proporcionando uma análise abrangente e detalhada. Para cada processo da base de teste, cada um dos 11 modelos fará uma predição binária: aqueles classificados como “sem risco” serão classificados como 0, enquanto aqueles classificados como “com risco” serão classificados como 1.

A partir de tais classificações individuais, é apropriado que se avalie qual rótulo foi o mais previsto pelos 11 modelos para cada processo. Como cada processo só pode ser classificado de forma binária (“sem risco” e “com risco”), sempre haverá concordância de ao menos 6 modelos, sendo tal aferição majoritária equivalente à escolha coletiva que representará cada processo, para fins de comparação com a performance individual de cada modelo.

Para possibilitar uma análise mais aprofundada do comportamento das classificações individuais e coletiva, convém decompor os processos da base de teste em subconjuntos formados de acordo com o grau de consenso, da seguinte forma:

1. **Subconjunto Consenso 11:** aqueles processos da base de teste em que todos os modelos classificaram da mesma forma;
2. **Subconjunto Consenso 10:** aqueles processos da base de teste em que 10 modelos classificaram de uma forma e 1 modelo classificou de forma distinta;
3. **Subconjunto Consenso 9:** aqueles processos da base de teste em que 9 modelos classificaram de uma forma e 2 modelos classificaram de forma distinta;

4. **Subconjunto Consenso 8:** aqueles processos da base de teste em que 8 modelos classificaram de uma forma e 3 modelos classificaram de forma distinta;
5. **Subconjunto Consenso 7:** aqueles processos da base de teste em que 7 modelos classificaram de uma forma e 4 modelos classificaram de forma distinta;
6. **Subconjunto Consenso 6:** aqueles processos da base de teste em que 6 modelos classificaram de uma forma e 5 modelos classificaram de forma distinta;

Espera-se, nesse contexto, que a performance se comporte de forma diretamente proporcional ao grau de concordância dos modelos entre si.

Para os fins da presente pesquisa, considera-se que a análise coletiva é crucial, pois permite compreender como a combinação de diferentes perspectivas técnicas pode ser comparada com as escolhas individuais de cada modelo. Além disso, suspeita-se que uma avaliação técnica composta de diversas abordagens possa minimizar potenciais deficiências específicas de um modelo ou outro modelo em relação a características específicas de dados. Assim, busca-se uma solução mais integrada e eficiente, capaz de capturar a complexidade dos dados de forma mais precisa e abrangente.

4.3 Métricas gerais de avaliação de desempenho

No que concerne à avaliação dos resultados atingidos pelos modelos e pelos comitês classificatórios, a matriz de confusão é uma ferramenta fundamental. A partir dela, visualiza-se o desempenho do algoritmo ao comparar as previsões feitas pelo modelo com os valores reais dos rótulos dos dados de teste. Tal matriz é uma tabela que resume os resultados das previsões em uma forma de contingência. A Tabela 3 apresenta tal informação:

Tabela 3 – Conceito de matriz de confusão

Real / Previsto	Previsto Positivo	Previsto Negativo
Real positivo	Verdadeiros Positivos (VP)	Falsos Negativos (FN)
Real negativo	Falsos Positivos (FP)	Verdadeiros Negativos (VN)

^a Fonte: Elaboração própria.

Cada célula da matriz de confusão tem um significado específico:

- **Verdadeiro Positivo (VP):** o número de casos positivos corretamente classificados como positivo pelo modelo;
- **Verdadeiro Negativo (VN):** o número de casos negativos corretamente classificados como negativos pelo modelo;
- **Falso Positivo (FP):** o número de casos negativos que foram incorretamente classificados como positivos pelo modelo;
- **Falso Negativo (FN):** o número de casos positivos que foram incorretamente classificados como negativos pelo modelo.

A partir dessa matriz, várias métricas de avaliação são usualmente utilizadas, a depender do objetivo da pesquisa, quais sejam:

1. **Acurácia (Accuracy):** A acurácia é a proporção de previsões corretas em relação ao total de previsões realizadas pelo modelo. É calculada como a razão entre a soma dos verdadeiros positivos (VP) e verdadeiros negativos (VN) pelo número total de observações (VP+VN+FP+FN). É uma métrica simples e intuitiva, sendo eficaz especialmente quando as classes estão balanceadas. Pode ser enganosa em conjuntos de dados desbalanceados, pois pode apresentar uma alta acurácia mesmo se o modelo estiver apenas prevendo a classe majoritária corretamente (FAWCETT, 2006);
2. **Precisão (Precision):** A precisão da classe positiva é a proporção de verdadeiros positivos (VP) entre todos os exemplos que foram classificados como positivos pelo modelo (VP+FP). A precisão da classe negativa, por sua vez, é a proporção de verdadeiros negativos (VN) entre todos os exemplos que foram classificados como negativos pelo modelo (VN+FN). Esta métrica é particularmente útil quando se quer minimizar o número de falsos negativos (POWERS, 2011);
3. **Revocação (Recall):** A revocação da classe positiva, ou sensibilidade, é a proporção de verdadeiros positivos (VP) entre todos os exemplos que realmente pertencem à classe positiva (VP+FN). A revocação da classe negativa, por sua vez, é a proporção

de verdadeiros negativos (VN) entre todos os exemplos que realmente pertencem à classe negativa (VN+FP); (POWERS, 2011)

4. **Medida F1 (F1 Score):** A medida F1 é a média harmônica entre precisão e revocação, fornecendo um único valor que balanceia ambas as métricas. É particularmente útil em cenários com classes desbalanceadas onde há necessidade de um equilíbrio entre precisão e revocação. A medida F1 é calculada como $2 * (\text{precisão} * \text{revocação}) / (\text{precisão} + \text{revocação})$ (SASAKI, 2007);
5. **Curva ROC (Receiver Operating Characteristic Curve):** A curva ROC é um gráfico que mostra a relação entre a taxa de verdadeiros positivos (VP) e a taxa de falsos positivos (FP) em vários limiares de classificação. É utilizada para avaliar a capacidade discriminativa de um modelo de classificação, permitindo visualizar o trade-off entre sensibilidade e especificidade (FAWCETT, 2006);
6. **Área Sob a Curva ROC (AUC-ROC):** A AUC-ROC é a área sob a curva ROC, que traça a taxa de verdadeiros positivos (VP) contra a taxa de falsos positivos (FP) em diferentes limiares de classificação. A AUC fornece uma medida agregada do desempenho em todos os limiares de classificação, sendo uma métrica robusta para avaliar a capacidade discriminativa de um modelo. Uma AUC próxima de 1 indica um modelo excelente, enquanto uma AUC próxima de 0.5 indica um modelo que não é melhor do que o acaso (FAWCETT, 2006);
7. **MCC (Matthews Correlation Coefficient):** O coeficiente de correlação de Matthews mede a qualidade das classificações binárias, levando em consideração verdadeiros e falsos positivos e negativos. MCC é uma métrica equilibrada que pode ser usada mesmo em conjuntos de dados desbalanceados, sendo calculada como $(VP * VN - FP * FN) / \sqrt{((VP + FP)(VP + FN)(VN + FP)(VN + FN))}$ (MATTHEWS, 1975);

Todavia, antes de definir qual métrica é adequada para a presente pesquisa, é importante destacar quais seriam os objetivos do TCE/PB ao implementar uma avaliação automatizada de risco baseada em aprendizagem supervisionada.

Considerando que aqueles processos que forem classificados como “sem risco” (previsão de classe 0) irão para o fim da fila de prioridade na análise pelos Auditores,

enquanto os processos classificados como “com risco” (previsão de classe 1) terão prioridade na análise pela área técnica, sob a ótica do objetivo do TCE/PB, é desejável que:

1. **Seja maximizada a quantidade de verdadeiros negativos (VN)**, pois o ganho da avaliação de risco automatizada é diretamente proporcional à quantidade de processos que não tem risco (classe 0) e que nesta classe são corretamente classificados, tornando mais eficiente o uso dos Auditores disponíveis;
2. **Seja minimizada a quantidade de falsos negativos (FN)**, ou seja, aqueles processos que têm risco (classe 1), mas cuja classificação foi indevidamente feita pelo modelo como sendo pertencente à classe “sem risco” (classe 0), ainda que tal ótica possa levar à um número maior de falsos positivos;

É importante destacar que, no contexto do TCE/PB, a quantidade de falsos positivos não seria tão preocupante. O problema maior para o Tribunal em adotar uma classificação automatizada não estaria em classificar algo sem risco como arriscado, mas sim em deixar de avaliar prioritariamente algo que realmente é arriscado, devido a uma classificação incorreta na classe sem risco (Falsos Negativos - FN).

Observa-se que a precisão da classe negativa (processos sem risco) é, entre as métricas comumente utilizadas, a que mais se alinha aos objetivos destacados. No entanto, como será abordado na próxima seção, existe uma métrica semelhante à precisão, porém mais intuitiva e adequada aos objetivos deste estudo.

4.4 Métrica de Benefício-Custo (BC)

No contexto do TCE/PB, é interessante que se utilize uma métrica de Benefício-Custo (BC), que, a um só tempo, avalie a relação entre a quantidade de registros verdadeiros negativos (VN) e a quantidade de registros falso negativos (FN) gerados nas classificações individuais e coletiva:

$$\text{BC da automatização} = \frac{\sum \text{Verdadeiros Negativos}}{\sum \text{Falsos Negativos}}$$

O benefício, representado pelo numerador da equação (aquilo que se quer maximizar), adviria da quantidade de processos que o modelo selecionado seria capaz de considerar corretamente como não prioritário (sem risco) e efetivamente colocar ao final da fila de análise. O custo, representado pelo denominador (aquilo que se quer minimizar), seriam os erros de se classificar como não arriscados processos que, em verdade, têm risco e devem ser prioritariamente analisados pelo Tribunal.

Por ser uma relação simples entre as quantidades de processos, considera-se que a interpretação desse indicador é mais intuitiva, oferecendo uma ideia clara da proporção entre acertos e erros que impactam os objetivos da automatização. Por exemplo, caso tal métrica tenha o valor de 2, isso indica que, para cada 2 processos de baixo risco que deixaram, corretamente, de ser priorizados, há um processo que foi incorretamente classificado como de baixo risco e deixou de ser priorizado.

Ao obtermos tal métrica para cada um dos modelos e para a escolha coletiva, consegue-se ponderar e ordenar, entre as diferentes opções, aqueles que mais agregam valor sob a ótica do Tribunal.

Além da definição do indicador BC para fins da automatização, é fundamental que se possa estimar o valor implícito de tal indicador na escolha aleatória atualmente realizada pela Auditoria, de forma que se possa estabelecer um *baseline* para a comparação dos resultados da automatização.

Neste norte, observa-se que a quantidade de processos com risco selecionados nas escolhas de processos feitas atualmente pela Auditoria corresponde a uma variável aleatória (VA) que segue a distribuição hipergeométrica. A hipergeométrica é uma distribuição de probabilidade discreta que descreve o número de sucessos obtidos em n seleções sem reposição de uma população finita de tamanho N , que contém exatamente K sucessos. Diferentemente da distribuição binomial, onde as retiradas são com reposição, na distribuição hipergeométrica cada retirada afeta a probabilidade das escolhas subsequentes (ROSS, 2014).

Uma VA com distribuição hipergeométrica possui alguns parâmetros, descritos, a seguir, de acordo com o contexto da presente pesquisa:

$$X \sim \text{Hipergeométrica}(n, K, N)$$

Onde,

N: quantidade total de processos em estoque para análise da Auditoria;

n: quantidade de seleções pela Auditoria;

K: total de processos com risco na base de seleção;

Neste tipo de VA, a probabilidade de k processos com risco serem selecionados em n retiradas sem reposição obedece à seguinte função de probabilidade:

$$P(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

Em n tentativas, a quantidade esperada de processos com risco a ser retirada é dada pela seguinte expressão:

$$E(X) = n \cdot \frac{K}{N}$$

Nesse contexto, para comparar com cada modelo que será testado e com a classificação coletiva, convém respondermos a seguinte pergunta: *Ao selecionarmos, ao acaso, n processos da base de seleção, sem reposição, qual é o quociente da quantidade de processos sem risco que acabam não sendo selecionados¹² pela quantidade de processos com risco que acabam não sendo selecionados¹³?*

$$\text{BC da escolha atual} = \frac{\text{Processos Sem Risco Não Selecionados (equivale a VN)}}{\text{Processos Com Risco Não Selecionados (equivale a FN)}}$$

¹² A quantidade de processos sem risco não selecionados equivale aos verdadeiros negativos da aplicação dos modelos, ou seja, aquilo que se quer maximizar, sendo o numerador da métrica de BC utilizada nesta pesquisa;

¹³ A quantidade de processos com risco não selecionados equivale aos falsos negativos da aplicação dos modelos, ou seja, aquilo que se quer minimizar, sendo o denominador da métrica de BC utilizada nesta pesquisa;

5 Dos resultados

Neste capítulo, são apresentadas estimativas do indicador BC no contexto da priorização aleatória realizada pela auditoria atualmente, da aplicação dos modelos propostos e da escolha coletiva. Por fim, são discutidas as considerações sobre a melhor solução testada e o progresso da automatização gerado com base na referida métrica de desempenho, comparando com a situação atual.

5.1 Do indicador BC da situação atual

Antes de avaliar o indicador BC na aplicação dos modelos, é fundamental que se possa estimar o valor de tal indicador implícito na escolha aleatória da ordem de processos atualmente realizada pela Auditoria.

Considerando um conjunto de dados de teste com: $N = 4.009$ processos¹; $K = 1.713$ processos com risco e n retiradas², calculamos as seguintes variáveis e métrica de BC para seleção aleatória de auditoria:

$$\text{Processos com risco selecionados (A)} = n \times K = 1.713 \times \frac{1.713}{4.009} = 732$$

$$\text{Processos sem risco selecionados (B)} = n - A = 1.713 - 732 = 981$$

$$\text{Processos com risco NÃO selecionados (C)} = K - A = 1.713 - 732 = 981$$

$$\text{Processos sem risco NÃO selecionados (D)} = N - K - B = 4.009 - 1.713 - 981 = 1.315$$

$$\text{BC da escolha aleatória} = \frac{\text{Processos Sem Risco Não Selecionados}}{\text{Processos Com Risco Não Selecionados}} = \frac{1.315}{981} = 1,34$$

¹ Conforme será destacado mais à frente, a base de testes corresponde a 25% do dataset total de 16.035 registros, perfazendo 4.009 processos. A base de treinamento, por sua vez, é composta de 75% dos registros, com 12.026 processos.

² Tendo em vista que há K processos com risco no estoque de processos, busca-se obter o BC implícito na escolha de igual número de processos ($n = K$) para análise priorizada pela Auditoria. No entanto, essa escolha é apenas uma referência para se demonstrar o cálculo realizado, tendo em vista que é possível comprovar que o valor do BC implícito independe do valor de n .

Portanto, implícito na escolha aleatória atualmente realizada pelo TCE/PB, há um BC de 1,34 processos sem risco não selecionados (benefício por eficiência) por cada processo com risco não selecionado (custo de não priorização).

5.2 Do indicador BC da automatização

O conjunto de dados, após todas as etapas da AED, é composto por 16.035 processos, dos quais 9.183 (57,3%) são processos sem risco (pertencentes à classe 0) e 6.852 (42,7%) são processos com risco (pertencentes à classe 1). Em outras palavras, os dados são balanceados (MIRZAEI, 2021).

Antes de treinar qualquer um dos modelos, o dataset foi dividido em uma base de treino e uma de teste utilizando a função `train_test_split` da biblioteca `scikit-learn` do Python. Essa divisão alocou, aleatoriamente, 75% dos dados para o treinamento (12.026 processos) e 25% para os testes (4.009 processos), mantendo-se as mesmas proporções de processos sem risco (57,3%) e com risco (42,7%) do dataset completo³. Contudo, uma única divisão dos dados não seria representativa do dataset completo, potencialmente favorecendo ou prejudicando determinados modelos, dependendo daquela distribuição específica dos dados que foi adotada.

É importante ressaltar que o momento em que a aposentadoria foi concedida não é relevante para a análise dos processos pela Auditoria, uma vez que todas as conferências que são feitas estão ligadas estritamente à regra legal que rege àquele benefício. Em outras palavras, aposentadorias concedidas em anos distintos, mas que sejam regidas pelo mesmo dispositivo normativo, terão análises equivalentes, dos mesmos requisitos legais. Por conta disso, o ano da concessão de cada processo não foi considerado na segmentação dos dados em bases de treino em teste

Para mitigar esse risco, foram realizadas 150 segmentações aleatórias dos dados, respeitando-se as proporções destacadas entre base de treinamento e teste e entre processos com risco e sem risco. A avaliação com diferentes composições de dados é fundamental para garantir que modelos empíricos não sejam testados apenas com padrões específicos, mas que também consigam generalizar bem em diversas situações. Modelos que consistentemente apresentam bom desempenho em múltiplas simulações têm maior probabilidade de

³ A proporção é mantida a partir da ativação do parâmetro `stratify`, dentro da função `train_test_split`.

generalizar bem para novos dados. **Dessa forma, as simulações permitem o cálculo de intervalos de confiança para as medidas da matriz de confusão e para o indicador BC, resultando em métricas mais confiáveis para a avaliação do desempenho tanto da escolha coletiva, quanto das individuais.**

Assim, todas as tabelas a seguir demonstradas são preenchidas com os intervalos de confiança calculados a partir das 150 iterações de partição dos dados, contemplando processos distintos de treinamento-teste-avaliação dos modelos. O procedimento ora adotado reflete a adoção da metodologia de validação cruzada, conhecida como K-Fold Cross Validation (LYASHENKO; JHA, 2024), com $K = 150$.

Nesse norte, os intervalos de confiança, além das margens de erro em valor absoluto, serão apresentados com o nível percentual do erro entre parênteses, da seguinte forma:

$$\text{Intervalo de Confiança} = \bar{X} \pm \text{Erro} \left(\left(\frac{\text{Erro}}{\bar{X}} - 1 \right) \times 100\% \right)$$

Com relação aos hiperparâmetros, há concretamente um *trade-off* entre a quantidade de modelos testados e o esforço computacional e de tempo demandado para a calibração ótima. Nesse contexto, considerou-se que usar um amplo substrato técnico derivado de uma maior gama de modelos justifica a adoção dos hiperparâmetros padrão da biblioteca *scikit-learn*. Portanto, os resultados demonstrados nesta seção provavelmente são passíveis de melhorias marginais tanto nas performances dos modelos quanto na escolha coletiva, mas esse potencial de melhoria não retira o valor dos resultados alcançados sem a otimização completa dos hiperparâmetros.

Em seguida, foram calculadas as matrizes de confusão de cada modelo e da escolha coletiva, bem como o indicador BC. Para determinar qual abordagem de automação, modelo individual ou escolha coletiva, é mais adequada ao objetivo desta pesquisa, é importante que cada iteração da base de teste seja segmentada em subconjuntos de processos, de acordo com o grau de consenso observado na classificação realizada pelos modelos. Nesse contexto, a matriz de confusão e o indicador BC da escolha coletiva é disposta na Tabela 4 seguir:

Tabela 4 – BC da escolha coletiva conforme grau de consenso dos modelos

Grau de con-senso	VN*	FP*	FN**	VP**	Qtd. Proc.	% da base de teste	BC
11 modelos	1039,46 ± 13,54 (1,30%)	75,31 ± 1,71 (2,27%)	343,45 ± 6,21 (1,81%)	237,20 ± 2,47 (1,04%)	1.695,4	42,30%	3,06 ± 0,06 (1,96%)
10 modelos	311,97 ± 10,74 (3,44%)	69,55 ± 1,59 (2,28%)	162,83 ± 4,05 (2,49%)	152,42 ± 2,44 (1,60%)	696,8	17,40%	1,92 ± 0,03 (1,70%)
9 modelos	177,18 ± 3,23 (1,83%)	63,92 ± 1,54 (2,40%)	112,75 ± 2,18 (1,93%)	114,88 ± 1,80 (1,57%)	468,7	11,70%	1,59 ± 0,03 (2,05%)
8 ou menos modelos	343,85 ± 3,86 (1,12%)	214,76 ± 2,76 (1,29%)	305,26 ± 3,10 (1,01%)	284,20 ± 3,48 (1,23%)	1.148,1	28,60%	1,13 ± 0,01 (1,22%)
Total					4.009	100,00%	

^a Fonte: Elaboração própria.

^b Nota: Intervalo de confiança a 95%.

^c * Qtd. de processos sem risco

^d ** Qtd. de processos com risco

Como se pode perceber, de 4.009 processos componentes da base de teste, em média, 1.695,4 (42,3% da base de teste) foram classificados com o mesmo rótulo por todos os modelos, gerando um indicador BC de 3,06, índice que equivale um avanço de 128,4% em relação à escolha aleatória atualmente realizada (1,34). Ou seja, boa parcela dos processos testados são classificados de forma unânime, o que indica que, a despeito das diferenças técnicas subjacentes aos modelos, há uma gama representativa de processos cuja escolha automatizada é significativamente superior ao processo atualmente adotado pela Auditoria. Como todos os modelos classificaram da mesma forma, a matriz de confusão para cada um dos modelos é a mesma e se encontra reproduzida na Tabela 5 a seguir:

Tabela 5 – BC da escolha de cada modelo para os processos em que 11 modelos concordaram

Modelo	VN*	FP*	FN**	VP**	Qtd. BC Proc.
Cada um dos modelos	1039,46 ± 13,54 (1,30%)	75,31 ± 1,71 (2,27%)	343,45 ± 6,21 (1,81%)	237,20 ± 2,47 (1,04%)	1.695,4 3,06 ± 0,06 (1,96%)

^a Fonte: Elaboração própria.

^b Nota: Intervalo de confiança a 95%.

^c * Qtd. de processos sem risco

^d ** Qtd. de processos com risco

Por conseguinte, para processos em que todos os 11 modelos concordaram, a escolha individual e a escolha coletiva oferecem o mesmo nível médio de BC (3,06).

Em relação aos 696,8 processos (17,4% da base de teste), em média, nos quais 10 modelos concordaram na classificação, o índice BC coletivo foi de 1,92, o que equivale a um avanço de 43,3% em relação à escolha aleatória atualmente realizada (1,34). Quanto a tais processos, apresenta-se, na Tabela 6 a seguir, o comportamento preditivo individual dos modelos que levaram à escolha coletiva destacada:

Tabela 6 – BC da escolha de cada modelo para os processos em que 10 modelos concordaram

Modelo	VN*	FP*	FN**	VP**	Qtd. BC Proc.
SVM	310,68 ± 10,77 (3,47%)	70,83 ± 1,55 (2,19%)	162,24 ± 4,10 (2,53%)	153,01 ± 2,45 (1,60%)	696,8 1,92 ± 0,03 (1,69%)
Logistic Regression	309,73 ± 10,74 (3,47%)	71,79 ± 1,64 (2,29%)	162,35 ± 4,12 (2,54%)	152,91 ± 2,44 (1,59%)	696,8 1,91 ± 0,03 (1,67%)

Tabela 6 – BC da escolha de cada modelo para os processos em que 10 modelos concordaram (*continuando*)

Modelo	VN*	FP*	FN**	VP**	Qtd. BC	Proc.
Random Forest	308,91 ± 10,78 (3,49%)	72,61 ± 1,57 (2,16%)	162,18 ± 4,09 (2,52%)	153,07 ± 2,49 (1,63%)	696,8	1,91 ± 0,03 (1,73%)
Bagging	308,93 ± 10,85 (3,51%)	72,59 ± 1,77 (2,44%)	163,25 ± 4,07 (2,49%)	152,01 ± 2,50 (1,64%)	696,8	1,90 ± 0,03 (1,71%)
Gradient Boost	310,35 ± 10,75 (3,46%)	71,16 ± 1,61 (2,26%)	163,82 ± 4,03 (2,46%)	151,43 ± 2,46 (1,62%)	696,8	1,90 ± 0,03 (1,69%)
MLPClassifier	275,55 ± 6,00 (2,18%)	105,97 ± 13,64 (12,87%)	149,03 ± 4,03 (2,71%)	166,22 ± 5,63 (3,38%)	696,8	1,87 ± 0,03 (1,84%)
Decision Tree	267,75 ± 11,37 (4,25%)	113,77 ± 2,28 (2,00%)	149,17 ± 4,27 (2,86%)	166,08 ± 2,46 (1,48%)	696,8	1,80 ± 0,03 (1,82%)
Extratree	295,18 ± 11,07 (3,75%)	86,33 ± 1,88 (2,18%)	175,53 ± 4,13 (2,35%)	139,72 ± 2,38 (1,71%)	696,8	1,68 ± 0,03 (1,82%)
KNN	219,36 ± 11,48 (5,24%)	162,15 ± 4,14 (2,55%)	132,19 ± 4,16 (3,14%)	183,07 ± 3,21 (1,76%)	696,8	1,66 ± 0,03 (2,01%)
ADA	310,09 ± 11,03 (3,56%)	71,43 ± 1,84 (2,58%)	190,87 ± 4,35 (2,28%)	124,39 ± 2,21 (1,77%)	696,8	1,63 ± 0,03 (1,82%)
Naive Bayes	272,70 ± 11,38 (4,17%)	108,81 ± 2,82 (2,59%)	170,13 ± 4,18 (2,45%)	145,13 ± 2,30 (1,58%)	696,8	1,60 ± 0,03 (1,93%)

Tabela 6 – BC da escolha de cada modelo para os processos em que 10 modelos concordaram (*continuando*)

Modelo	VN*	FP*	FN**	VP**	Qtd. BC Proc.
--------	-----	-----	------	------	------------------

^a Fonte: Elaboração própria.

^b Nota: Intervalo de confiança a 95%.

^c * Qtd. de processos sem risco

^d ** Qtd. de processos com risco

Para tal subconjunto, observa-se que a escolha coletiva apresenta uma relação BC ($1,92 \pm 0,03$) equivalente estatisticamente à performance dos modelos SVM ($1,92 \pm 0,03$), Regressão Logística ($1,91 \pm 0,03$), Random Forest ($1,91 \pm 0,03$), Bagging ($1,90 \pm 0,03$) e Gradient Boost ($1,90 \pm 0,03$).

Em relação aos 468,7 processos (11,7% da base de teste), em média, nos quais 9 modelos concordaram na classificação, o índice BC coletivo foi de 1,59, o que equivale a um avanço de 17,9% em relação à escolha aleatória (1,34). Quanto a tais processos, apresenta-se, na Tabela 7 a seguir, o comportamento preditivo individual dos modelos que levaram à escolha coletiva destacada:

Tabela 7 – BC da escolha de cada modelo para os processos em que 9 modelos concordaram

Modelo	VN*	FP*	FN**	VP**	Qtd. Benefício Total de / Custo Proc.
SVM	173,11 ±	67,99 ±	110,43 ±	117,21 ±	468,7 1,58 ± 0,03 (2,11%)
	3,24	1,65	2,14	1,85	
	(1,87%)	(2,43%)	(1,93%)	(1,57%)	
Bagging	168,36 ±	72,74 ±	109,65 ±	117,99 ±	468,7 1,55 ± 0,03 (2,09%)
	3,20	1,72	2,17	1,96	
	(1,90%)	(2,36%)	(1,98%)	(1,66%)	

Tabela 7 – BC da escolha de cada modelo para os processos em que 9 modelos concordaram
(continuando)

Modelo	VN*	FP*	FN**	VP**	Qtd. Total de Proc.	Beneficio / Custo
Logistic Regression	171,33 ±	69,77 ±	113,05 ±	114,59 ±	468,7	1,53 ±
	3,31	1,57	2,19	1,88		0,03
	(1,93%)	(2,25%)	(1,94%)	(1,64%)		(2,05%)
Random Forest	164,69 ±	76,41 ±	108,59 ±	119,05 ±	468,7	1,53 ±
	3,36	1,86	2,27	1,91		0,03
	(2,04%)	(2,44%)	(2,09%)	(1,60%)		(2,15%)
MLPClassifier	147,09 ±	94,01 ±	100,83 ±	126,80 ±	468,7	1,50 ±
	4,32	4,77	3,64	3,59		0,05
	(2,94%)	(5,07%)	(3,61%)	(2,83%)		(3,31%)
Gradient Boost	173,08 ±	68,02 ±	120,38 ±	107,25 ±	468,7	1,45 ±
	3,25	1,66	2,36	1,76		0,03
	(1,88%)	(2,44%)	(1,96%)	(1,65%)		(2,03%)
Decision Tree	135,96 ±	105,14 ±	95,12 ±	132,51 ±	468,7	1,44 ±
	2,96	1,99	2,00	2,25		0,03
	(2,18%)	(1,89%)	(2,10%)	(1,70%)		(2,29%)
Extratree***	160,67 ±	80,43 ±	123,29 ±	104,34 ±	468,7	1,31 ±
	3,21	2,06	2,25	1,98		0,03
	(2,00%)	(2,57%)	(1,82%)	(1,90%)		(2,08%)
KNN***	133,07 ±	108,03 ±	105,35 ±	122,28 ±	468,7	1,27 ±
	2,75	2,28	2,06	2,05		0,03
	(2,07%)	(2,11%)	(1,95%)	(1,68%)		(2,29%)
ADA***	160,95 ±	80,15 ±	139,04 ±	88,59 ±	468,7	1,16 ±
	3,20	2,08	2,49	1,84		0,02
	(1,99%)	(2,60%)	(1,79%)	(2,08%)		(1,91%)

Tabela 7 – BC da escolha de cada modelo para os processos em que 9 modelos concordaram
(continuando)

Modelo	VN*	FP*	FN**	VP**	Qtd. Total de Proc.	Benefício / Custo
Naive Bayes***	134,15 ± 2,82 (2,11%)	106,95 ± 2,33 (2,18%)	118,81 ± 2,05 (1,73%)	108,82 ± 2,04 (1,88%)	468,7	1,14 ± 0,02 (2,18%)

^a Fonte: Elaboração própria.

^b Nota: Intervalo de confiança a 95%.

^c * Qtd. de processos sem risco

^d ** Qtd. de processos com risco

^e *** Modelos com BC inferior à escolha aleatória (1,34) para tal subconjunto.

Nota-se que a escolha coletiva para processos em que 9 modelos concordaram apresenta um BC ($1,59 \pm 0,03$) estatisticamente equivalente aos apresentados pelos modelos SVM ($1,58 \pm 0,03$) e Bagging ($1,55 \pm 0,03$).

Em relação aos demais 1.148,1 processos em que 8 ou menos modelos concordaram, o índice de BC coletivo foi de 1,13, o que equivale a uma perda média de 15,7% em relação à escolha aleatória. Quanto a tais processos, apresenta-se, na Tabela 8 a seguir, o comportamento preditivo individual dos modelos que levaram à escolha coletiva destacada:

Tabela 8 – BC da escolha de cada modelo para os processos em que 8 ou menos modelos concordaram

Modelo	VN*	FP*	FN**	VP**	Qtd. Total de Proc.	Benefício / Custo
Bagging***	319,03 ± 3,92 (1,23%)	239,59 ± 3,16 (1,32%)	257,09 ± 3,35 (1,30%)	332,37 ± 3,70 (1,11%)	1.148,1	1,25 ± 0,02 (1,31%)

Tabela 8 – BC da escolha de cada modelo para os processos em que 8 ou menos modelos concordaram (*continuando*)

Modelo	VN*	FP*	FN**	VP**	Qtd. Total de Proc.	Beneficio / Custo
Decision Tree***	260,14 ± 3,54 (1,36%)	298,47 ± 3,57 (1,20%)	222,89 ± 3,09 (1,39%)	366,57 ± 3,76 (1,02%)	1.148,1	1,17 ± 0,02 (1,43%)
Random Forest***	290,55 ± 4,10 (1,41%)	268,07 ± 3,04 (1,13%)	250,33 ± 3,67 (1,46%)	339,13 ± 3,30 (0,97%)	1.148,1	1,16 ± 0,01 (1,25%)
SVM***	342,89 ± 4,65 (1,35%)	215,73 ± 2,96 (1,37%)	308,47 ± 4,08 (1,32%)	280,99 ± 3,41 (1,21%)	1.148,1	1,11 ± 0,01 (1,22%)
KNN***	269,22 ± 3,57 (1,32%)	289,39 ± 3,47 (1,20%)	252,74 ± 3,25 (1,28%)	336,72 ± 3,65 (1,08%)	1.148,1	1,07 ± 0,01 (1,36%)
Extratree***	318,87 ± 4,20 (1,32%)	239,75 ± 2,71 (1,13%)	300,37 ± 3,79 (1,26%)	289,09 ± 3,17 (1,10%)	1.148,1	1,06 ± 0,01 (1,19%)
Logistic Regression***	326,91 ± 4,19 (1,28%)	231,71 ± 2,35 (1,01%)	344,81 ± 3,57 (1,04%)	244,65 ± 2,89 (1,18%)	1.148,1	0,95 ± 0,01 (1,15%)
MLPClassifier***	278,66 ± 8,54 (3,07%)	279,95 ± 10,46 (3,74%)	296,73 ± 9,54 (3,21%)	292,73 ± 11,26 (3,85%)	1.148,1	0,94 ± 0,02 (1,66%)
Gradient Boost***	316,89 ± 4,24 (1,34%)	241,73 ± 3,01 (1,25%)	351,12 ± 4,03 (1,15%)	238,34 ± 2,93 (1,23%)	1.148,1	0,90 ± 0,01 (1,15%)

Tabela 8 – BC da escolha de cada modelo para os processos em que 8 ou menos modelos concordaram (*continuando*)

Modelo	VN*	FP*	FN**	VP**	Qtd. Total de Proc.	Benefício / Custo
Naive Bayes***	266,80 ± 3,51 (1,32%)	291,81 ± 3,38 (1,16%)	306,07 ± 3,77 (1,23%)	283,39 ± 2,91 (1,03%)	1.148,1	0,87 ± 0,01 (1,14%)
ADA***	322,73 ± 4,20 (1,30%)	235,88 ± 3,69 (1,56%)	374,62 ± 4,33 (1,16%)	214,84 ± 3,14 (1,46%)	1.148,1	0,86 ± 0,01 (1,01%)

^a Fonte: Elaboração própria.

^b Nota: Intervalo de confiança a 95%.

^c * Qtd. de processos sem risco

^d ** Qtd. de processos com risco

^e *** Modelos com BC inferior à escolha aleatória (1,34) para tal subconjunto.

Observa-se que tanto a escolha coletiva quanto a melhor escolha individual para esses processos geram índices BC consideravelmente inferiores à escolha aleatória.

À vista de todas essas considerações, no contexto da escolha coletiva, observa-se que prevalece a automatização nos subconjuntos em que 9 ou mais modelos concordaram. Nos casos em que 8 ou menos modelos concordam, a escolha aleatória é superior. Nesse norte, o BC resultante⁴ da estratégia mista entre escolha coletiva e escolha puramente aleatória corresponde a 2,197, conforme detalhado na Tabela 9 seguir:

⁴ O BC resultante é calculado a partir da média de tal indicador para cada segmento ponderada pela quantidade respectiva de processos.

Tabela 9 – BC resultante da estratégia mista entre escolha coletiva e a escolha puramente aleatória

Grau de consenso	Escolha definida	Qtd. Proc.	Composição (%)	Composição Acumulada (%)	BC
11 modelos	Coletiva	1.695,4	42,30%	42,30%	3,06 ± 0,06 (1,96%)
10 modelos	Coletiva	696,8	17,40%	59,70%	1,92 ± 0,03 (1,70%)
9 modelos	Coletiva	468,7	11,70%	71,40%	1,59 ± 0,03 (2,05%)
8 ou menos modelos	Aleatória	1.148,1	28,60%	100,00%	1.34
Total		4.009	100,00%	BC Resultante →	2,197

^a Fonte: Elaboração própria.

^b Nota: Intervalo de confiança a 95%.

No que concerne à aplicação individual de um modelo, a partir das tabelas 6, 7 e 8, podemos perceber que o modelo SVM é aquele com melhor rendimento nos processos em que houve concordância de ao menos 9 modelos (processos esses em que as melhores escolhas individuais são superiores à escolha puramente aleatória). Quanto aos processos com consenso de 8 ou menos modelos, observa-se que a escolha aleatória é superior. Nesse norte, o BC resultante da estratégia mista entre modelo SVM e escolha puramente aleatória corresponde ao valor de 2,196, conforme detalhado na Tabela 10 a seguir:

Tabela 10 – BC resultante da estratégia mista entre a melhor escolha individual agregada e a escolha puramente aleatória

Grau de consenso	Escolha definida	Qtd. Proc.	Composição (%)	Composição Acumulada (%)	BC
11 modelos	SVM	1.695,4	42,30%	42,30%	3,06 ± 0,06 (1,96%)
10 modelos	SVM	696,8	17,40%	59,70%	1,92 ± 0,03 (1,69%)
9 modelos	SVM	468,7	11,70%	71,40%	1,58 ± 0,03 (2,11%)
8 ou menos modelos	Aleatória	1.148,1	28,60%	100,00%	1.34
Total		4.009	100,00%	BC Resultante →	2,196

^a Fonte: Elaboração própria.

^b Nota: Intervalo de confiança a 95%.

Do exposto, observa-se que a estratégia mista utilizando o melhor modelo (SVM - 2,196) é equivalente estatisticamente à estratégia mista utilizando a escolha coletiva (2,197). Tendo em vista que a avaliação coletiva tem o potencial de compensar eventuais deficiências de modelos individuais, considera-se que seria mais adequada para a avaliação de risco de processos de aposentadoria, com um incremento estimado em relação à situação atual de 64,0%:

$$\text{Ganho em Relação à situação atual} = \frac{\text{BC da Escolha Coletiva}}{\text{BC da Situação Atual}} = \frac{2,197}{1,340} - 1 = 64,0\%$$

5.3 Da importância dos atributos

Com intuito de propiciar ao leitor noções da influência relativa das variáveis explicativas (atributos) nas previsões da automatização, utilizou-se o modelo SVM, a partir de uma segmentação do *dataset* em conjuntos de treino e teste, por ser um modelo linear cujos coeficientes podem ser facilmente extraídos em código Python e por ter se destacado nos resultados individuais apresentados em relação aos demais modelos.

Para avaliar a influência relativa de cada variável preditora no modelo, foi realizada uma análise das importâncias agregadas por variável explicativa. No presente estudo, como tais variáveis são todas categóricas, tornou-se necessário convertê-las em tantas variáveis *dummy* quanto seus valores únicos componentes deduzidos de um, por meio do comando *OneHotEncoding*. Nesse contexto, a Tabela 11 apresenta tal as quantidades de variáveis *dummy* para cada campo originário:

Tabela 11 – Quantidade de variáveis dummy por campo categórico

Campo	Qtd. de Categorias	Qtd. de Dummies
jurisdicionado	6	5
ano	8	7
orgao	24	23
carreira	12	11
cargo	37	36
flag_aposesp	2	1
meio_publi	4	3
regra_apos	6	5
qtd_vinculos	5	4
qtd_vinculos_pub	4	3
qtd_vinculos_priv	4	3
qtd_vinculos_mesmo_cargo	3	2
qtd_vinculos_mesma_carreira	3	2
p_vencimento	2	1
p_ats	2	1

Tabela 11 – Quantidade de variáveis dummy por campo categórico (*continued*)

Campo	Qtd. de Categorias	Qtd. de Dummies
p_proventos	2	1
p_gratificacao	2	1
p_vantpessoal	2	1
p_insalubridade	2	1
p_adcqualificacao	2	1
p_representacao	2	1
p_antecipacao	2	1
p_indenizatorio	2	1
ind_sm	2	1
rotulo	2	1

^a Fonte: Elaboração própria.

Assim, para calcular a importância geral de cada campo categórico no modelo, foi necessário consolidar as contribuições de todas as variáveis dummy derivadas de cada campo.

O processo para a análise agregada seguiu os seguintes passos:

- (a) **Coefficientes das Variáveis:** Durante o treinamento do modelo supervisionado (neste caso, *LinearSVC* com *kernel* linear), cada variável recebeu um coeficiente associado, representando seu peso no modelo. Esses coeficientes indicam o impacto relativo das variáveis nas decisões do modelo;
- (b) **Transformação em Valores Absolutos:** Como os coeficientes podem assumir valores positivos ou negativos, foi utilizada a magnitude (valor absoluto) dos coeficientes para capturar a relevância total de cada variável, independentemente da direção de sua influência;
- (c) **Agrupamento por Campo Original:** As variáveis derivadas de um mesmo campo categórico foram agrupadas com base no nome do campo original, identificado antes da aplicação do *OneHotEncoding*. Para cada campo, as importâncias (valores

absolutos dos coeficientes) de todas as variáveis *dummy* geradas foram somadas, resultando em uma métrica consolidada de importância para o campo original;

- (d) **Ordenação e Apresentação:** As importâncias consolidadas foram ordenadas de forma decrescente, permitindo identificar os campos preditores mais relevantes no contexto do modelo treinado.

Essa abordagem permite uma interpretação global do impacto de cada variável preditora, mesmo em casos onde o pré-processamento resulta em um grande número de variáveis derivadas.

A Tabela 12 apresenta os resultados da análise de importância das variáveis preditoras, agrupadas por campo original, destacando aqueles com maior relevância no modelo:

Tabela 12 – Grau de importância de cada atributo de treinamento

Campo Originário	Importância
orgao	5,01
cargo	3,76
carreira	1,49
ano	1,35
jurisdicionado	1,18
regra_apos	1,07
p_indenizatorio	0,87
meio_publico	0,37
flag_aposesp	0,26
qtd_vinculos_mesmo_cargo	0,22
qtd_vinculos	0,21
qtd_vinculos_priv	0,20
qtd_vinculos_mesma_carreira	0,20
p_proventos	0,19
p_atc	0,16
ind_sm	0,15

Tabela 12 – Grau de importância de cada atributo de treinamento (*continued*)

Campo Originário	Importância
p_adcqualificacao	0,13
qtd_vinculos_pub	0,12
p_antecipacao	0,11
p_vencimento	0,10
p_representacao	0,09
p_insalubridade	0,08
p_gratificacao	0,05
p_vantpessoal	0,05

^a Fonte: Elaboração própria.

Da tabela acima, percebe-se que a soma dos valores absolutos das variáveis *dummy* decorrentes de cada campo indica que os atributos 'orgao', 'cargo' e 'carreira' foram os atributos com maior influência na predição do modelo SVM. É interessante ressaltar que tais campos são alguns dos que precisaram de grande número de transformações e de formação de categorias para propiciar padrões para o treinamentos dos modelos e, ainda assim, ainda há muitos registros no *dataset* em que não se conseguiu identificar conteúdo inteligível. Portanto, uma normatização e padronização do preenchimento a ser feito pelos institutos de previdência pode ser muito benéfica para a automatização, com potencial para melhorar ainda mais os resultados observados.

Por fim, convém destacar a maior relevância, em relação a outras rubricas, de parcelas indenizatórias nos proventos dos aposentados como indício que há irregularidade na concessão do benefício.

6 Considerações Finais

Este trabalho teve como objetivo central comprovar a viabilidade da identificação do risco intrínseco de irregularidades em processos de aposentadoria utilizando modelos de classificação automatizada por meio de *machine learning*. Especificamente, busca-se contribuir para a otimização da gestão de risco do TCE/PB, para o uso mais racional da mão de obra técnica e para a expansão da área de atuação do setor de inteligência da Corte de Contas Paraibana.

Para atingir esses objetivos, a partir de dados estruturados de aposentadorias enviados pelos institutos de previdência ao sistema TRAMITA do TCE/PB, criou-se um rótulo binário para cada processo de concessão. Esse rótulo assumiu o valor 0 nos processos em que a Auditoria do TCE/PB precisou formalizar apenas um relatório técnico, sem identificar irregularidades, e o valor 1 para os processos que exigiram mais de uma intervenção técnica, com identificação de irregularidades na concessão e, conseqüentemente, maior risco.

Em seguida, foi realizada uma ampla análise exploratória de dados, que revelou a existência de uma quantidade relevante de dados para o uso de técnicas de *machine learning*, mas também identificou que muitos atributos importantes para a aprendizagem são de preenchimento livre, sem categorização ou padronização. Isso exigiu um grande esforço de pré-processamento dos dados textuais para viabilizar a captura de padrões consistentes e úteis à identificação de risco, consolidando a um conjunto de dados com 16.035 processos, divididos em 75% (12.026) para a base de treinamento e 25% (4.009) para a base de teste.

Considerando o contexto específico do TCE/PB, onde a aprendizagem supervisionada pode ser benéfica pela quantidade de processos sem risco corretamente classificados e custosa pela quantidade de processos com risco indevidamente classificados como sem risco, foi definida uma métrica de avaliação baseada no BC da automatização, obtida a partir da matriz de confusão. Esse indicador é calculado pelo quociente entre os verdadeiros negativos (processos sem risco corretamente classificados) e os falsos negativos (processos com risco indevidamente classificados como sem risco).

Foram treinados 11 modelos com embasamentos técnicos e teóricos diversos. Além da análise individual de cada modelo, para aumentar a confiabilidade da análise automatizada, foi avaliada a escolha coletiva derivada das classificações dos modelos, segmentando-se a base de teste de acordo com o grau de consenso obtido. Ademais, para evitar vieses nos achados, foram realizadas 150 simulações de treinamento e avaliação de resultados, apresentando-se intervalos com 95% de confiança da performance dessas iterações.

Nesse contexto, observou-se que, para os processos em que houve concordância na classificação entre 9 ou mais modelos, a escolha coletiva foi superior à escolha aleatória e estatisticamente equivalente ao melhor modelo individualmente considerado. Todavia, para os processos em que houve consenso de classificação de 8 ou menos modelos, tanto a escolha coletiva quanto os modelos individualmente considerados apresentaram um BC médio inferior à escolha aleatória atualmente realizada pela Auditoria. Assim, ao considerar a escolha coletiva para os processos em que houve consenso de ao menos 9 modelos e a aleatória para os demais processos, obteve-se um BC resultante médio de 2,197, 64,0% superior ao processo atual de escolha aleatória pura (1,34), e equivalente ao propiciado pelo melhor modelo (2,196, SVM). Nesse contexto, considera-se que a escolha coletiva é mais robusta, uma vez que tem potencial de compensar eventuais deficiências apresentadas por modelos individuais.

Portanto, ainda que se considerem as deficiências da base de dados, a avaliação automatizada propiciou, além da possibilidade de otimização dos recursos humanos envolvidos, uma ordenação de análise de processos com ganhos expressivos em relação à seleção atualmente conduzida pelo TCE/PB, com potencial de uso no curto prazo para processos de aposentadoria regidos por regras anteriores à Reforma da Previdência ocorrida em 2019. Para tanto, é fundamental que o Tribunal realize um esforço de categorização dos principais campos e de regulamentação do preenchimento a ser feito pelos institutos de previdência, de forma a possibilitar a identificação de padrões úteis e consistentes para melhor performance dos modelos.

Nesse contexto, esta pesquisa demonstra a viabilidade de melhorias na eficiência da análise de aposentadorias por parte do Tribunal de Contas da Paraíba, ao proporcionar uma priorização mais precisa na análise dos processos. Ademais, a possibilidade de identificar, de forma automatizada, potenciais irregularidades alinha-se à crescente necessidade de assegurar que os recursos públicos sejam aplicados de forma adequada e apenas onde

são devidos, especialmente considerando o impacto crescente desse tipo de benefício previdenciário nos orçamentos do Estado da Paraíba e de seus municípios, em face do acelerado envelhecimento populacional.

Em evoluções do presente estudo, não se pode olvidar, ainda, a possibilidade de incorporação de outras fontes de dados relevantes que contemplem aspectos importantes relacionados à concessão de aposentadorias, o que poderia aumentar ainda mais o poder explicativo dos modelos de *machine learning*.

Por fim, é imperioso que sejam respondidas as questões de pesquisa apresentadas no item 1.1 desta dissertação.

Quanto à QP1¹, não obstante haja deficiências na qualidade dos dados fornecidos pelos institutos de previdência, atividades de pré-processamento permitiram a realização de classificação a *priori* de risco, tanto pelos modelos individualmente considerados quanto por avaliação coletiva.

No que concerne à QP2², a classificação coletiva, obtida pela votação majoritária entre os 11 modelos contemplados na pesquisa, propicia uma relação de Benefício-Custo superior à escolha aleatória atualmente realizada pelo TCE/PB para aqueles processos em que há consenso de ao menos 9 modelos, passível de ser aplicado na priorização dos processos de aposentadoria.

¹ **QP1:** *Os dados estruturados relativos aos processos de aposentadoria enviados ao TCE/PB pelos institutos de previdência são suficientes para uma classificação, a priori, confiável de risco?*

² **QP2:** *Há um modelo ou agrupamento de modelos de aprendizagem supervisionada mais adequado ao objetivo da pesquisa?*

Referências

- ASOMURA, I.; IJIMA, R.; MORI, T. Automating the detection of fraudulent activities in online banking service. *Journal of Information Processing*, v. 31, p. 643–653, 2023. ISSN 1882-6652. Disponível em: <https://www.jstage.jst.go.jp/article/ipsjjip/31/0/31_643/article/-char/en>. Citado na página 16.
- BIAU, G.; DEVROYE, L. *Lectures on the Nearest Neighbor Method*. [S.l.]: Springer, 2015. (Springer Series in the Data Sciences). Citado na página 28.
- BIAU, G.; SCORNET, E. A random forest guided tour. *Online*, 2016. [Citado em: 14 de mar. de 2024]. Disponível em: <<https://doi.org/10.1007/s11749-016-0481-7>>. Citado na página 29.
- BRANDÃO, M. A. et al. Plus: A semi-automated pipeline for fraud detection in public bids. *Digital Government: Research and Practice*, v. 5, n. 1, p. Article 5 (March 2024), 16 pages, 2024. Disponível em: <<https://doi-org.ez15.periodicos.capes.gov.br/10.1145/3616396>>. Citado na página 18.
- BREIMAN, L. Bagging predictors. *Machine Learning*, v. 24, n. 2, p. 123–140, 1996. Citado na página 29.
- COSTA, M. B.; BASTOS, P. R. L. Alice, monica, adele, sofia, carina e Ágata: o uso da inteligência artificial pelo tribunal de contas da união. *Revista do Tribunal de Contas do Estado de Goiás*, 2020. Citado na página 4.
- DOMASHOVA, J.; KRIPAK, E. Identification of non-typical international transactions on bank cards of individuals using machine learning methods. *Procedia Computer Science*, v. 190, p. 178–183, 2021. ISSN 1877-0509. Disponível em: <<https://doi.org/10.1016/j.procs.2021.06.023>>. Citado na página 16.
- ELSADIG, M. et al. Intelligent deep machine learning cyber phishing url detection based on bert features extraction. *Electronics*, v. 11, p. 3647, 2022. Disponível em: <<https://doi.org/10.3390/electronics11223647>>. Citado na página 17.
- FAKIHA, B. Forensic credit card fraud detection using deep neural network. *Journal Of Southwest Jiaotong University*, 2023. ISSN 0258-2724. Citado na página 16.
- FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006. Disponível em: <<https://doi.org/10.1016/j.patrec.2005.10.010>>. Citado 2 vezes nas páginas 32 e 33.

FREUND, Y.; SCHAPIRE, R. E. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, v. 14, n. 5, p. 771–780, 1999. (Em japonês, tradução de Naoki Abe). Citado na página 30.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, v. 29, n. 5, p. 1189–1232, 2001. Citado na página 29.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. *Machine Learning*, v. 63, p. 3–42, 2006. Disponível em: <<https://doi.org/10.1007/s10994-006-6226-1>>. Citado na página 29.

HAMZA, C. et al. Semi-supervised method to detect fraudulent transactions and identify fraud types while minimizing mounting costs. *International Journal of Advanced Computer Science and Applications (IJACSA)*, v. 14, n. 2, 2023. Disponível em: <<http://dx.doi.org/10.14569/IJACSA.2023.0140298>>. Citado na página 16.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning*. 2nd. ed. New York: Springer, 2009. Citado na página 29.

HENRIQUE, B. M.; SOBREIRO, V. A.; KIMURA, H. Contracting in brazilian public administration: A machine learning approach. *Expert Systems*, v. 37, p. e12550, 2020. Disponível em: <<https://doi.org/10.1111/exsy.12550>>. Citado na página 17.

HERLAND, M.; BAUDER, R. A.; KHOSHGOFTAAR, T. M. The effects of class rarity on the evaluation of supervised healthcare fraud detection models. *Journal of Big Data*, v. 6, p. 21, 2019. Disponível em: <<https://doi.org/10.1186/s40537-019-0181-8>>. Citado na página 16.

LABANCA, D. et al. Amaretto: An active learning framework for money laundering detection. *IEEE Access*, v. 10, p. 41720–41739, 2022. Citado na página 16.

LOKANAN, M. E. Predicting mobile money transaction fraud using machine learning algorithms. *Applied AI Letters*, v. 4, p. e85, 2023. Disponível em: <<https://doi.org/10.1002/ail2.85>>. Citado na página 16.

LOKANAN, M. E.; SHARMA, K. Fraud prediction using machine learning: The case of investment advisors in canada. *Machine Learning with Applications*, v. 8, 2022. Disponível em: <<https://doi.org/10.1016/j.mlwa.2022.100269>>. Citado na página 16.

LYASHENKO, V.; JHA, A. *Cross-Validation in Machine Learning: How to Do It Right*. 2024. Online. [Citado em: 19 de ago. de 2024]. Disponível em: <<https://neptune.ai/blog/cross-validation-in-machine-learning-how-to-do-it-right>>. Citado na página 39.

LYRA, M. S. et al. Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods. *Applied Network Science*, v. 7, p. 83, 2022. Disponível em: <<https://doi.org/10.1007/s41109-022-00523-6>>. Citado na página 18.

MATTHEWS, B. W. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, v. 405, n. 2, p. 442–451, 1975. Disponível em: <[https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9)>. Citado na página 33.

MIRZAEI, A. *Imbalanced Classification*. 2021. LinkedIn. Disponível em: <<https://www.linkedin.com/pulse/imbalanced-classification-ali-mirzaei>>. Citado na página 38.

NICHOLLS, J.; KUPPA, A.; LE-KHAC, N.-A. Financial cybercrime: A comprehensive survey of deep learning approaches to tackle the evolving financial crime landscape. *IEEE Access*, v. 9, p. 163965–163986, 2021. Citado na página 17.

PAMPEL, F. C. *Logistic Regression: A Primer*. [S.l.]: SAGE Publications, Inc., 2020. (Quantitative Applications in the Social Sciences). Citado na página 28.

POWERS, D. M. W. Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, v. 2, n. 1, p. 37–63, 2011. Disponível em: <https://www.researchgate.net/publication/275507764_Evaluation_From_Precision_Recall_and_F-Measure_to_ROC_Informedness_Markedness_and_Correlation>. Citado 2 vezes nas páginas 32 e 33.

QUINLAN, J. R. Induction of decision trees. *Online*, 2024. [Citado em: 14 de mar. de 2024]. Disponível em: <<https://doi.org/10.1007/BF00116251>>. Citado na página 28.

RANDHAWA, K. et al. Credit card fraud detection using adaboost and majority voting. *IEEE Access*, v. 6, p. 14277–14284, 2018. Citado na página 16.

RISH, I. An empirical study of the naive bayes classifier. In: *IJCAI 2001 Work Empir Methods Artif Intell*. [S.l.: s.n.], 2001. v. 3. Citado na página 28.

RODRÍGUEZ, M. J. G. et al. Collusion detection in public procurement auctions with machine learning algorithms. *Automation in Construction*, v. 133, p. 104047, 2022. ISSN 0926-5805. Disponível em: <<https://doi.org/10.1016/j.autcon.2021.104047>>. Citado na página 18.

ROSS, S. M. *Introduction to Probability and Statistics for Engineers and Scientists*. 5th. ed. [S.l.]: Academic Press, 2014. Citado na página 35.

SASAKI, Y. The truth of the f-measure. *Teach Tutor Mater*, v. 1, n. 5, p. 1–5, 2007. Disponível em: <<https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf>>. Citado na página 33.

SCHOLKOPF, B.; SMOLA, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. [S.l.]: MIT Press, 2018. (Adaptive Computation and Machine Learning series). Citado na página 28.

STEFÁNSSON, H. et al. *Detecting potential money laundering addresses in the Bitcoin blockchain using unsupervised machine learning*. [S.l.: s.n.], 2022. Citado na página 16.

WEN, S. et al. Analysis of financial fraud based on manager knowledge graph. *Procedia Computer Science*, v. 199, p. 773–779, 2022. ISSN 1877-0509. Disponível em: <<https://doi.org/10.1016/j.procs.2022.01.096>>. Citado na página 16.

Anexos

ANEXO A – Protocolo de Revisão da Literatura

Conforme discutido no capítulo de revisão da literatura (Capítulo 3), o levantamento foi realizado tendo por amparo este protocolo, onde se buscou obter os trabalhos mais relevantes sobre a aprendizagem de máquina e/ou algoritmos preditivos e que fossem aplicados no contexto da Administração Pública, notadamente nos aspectos de avaliação de risco, controle e tomada de decisão, aumento da produtividade e otimização da eficiência. Com esse objetivo, a busca por trabalhos relacionados pautou-se pelas seguintes questões de pesquisa:

- **QP1:** é possível identificar contribuições científicas que usem técnicas computacionais votadas ao gerenciamento de risco, tomada de decisão e predição da conformidade de procedimentos?;
- **QP2:** essas contribuições foram identificadas em quais contextos (Governamental, Privado)?;
- **QP3:** quais as metodologias atualmente utilizadas com a finalidade de predição do risco desses processos; e
- **QP4:** a aprendizagem de máquina, como metodologia de análise, tem sido utilizada no processo de predição de risco? Ela pode auxiliar no processo de tomada de decisão?

De forma a executar a pesquisa pelos trabalhos relacionados na área, foi feita uma exploração automática utilizando-se o engenho de busca oferecido pela CAPES, disponível no endereço eletrônico <https://www-periodicos-capes-gov-br.ez15.periodicos.capes.gov.br/>, que cobre um vasto campo do conhecimento humano, e sendo integrado pelos mais diversos repositórios, como Scopus, IEEE Xplore e Web of Science.

Nesse contexto, foram definidas as palavras-chave que seriam empregadas nas pesquisas e que são utilizadas pelo engenho para retornar os trabalhos que contenham

os termos, de acordo com as opções definidas no motor de busca. Neste, por sua vez, os filtros foram considerados de maneira a retornar os trabalhos que contivessem os termos de pesquisa em qualquer tópico do artigo científico.

Na Tabela 13, é possível visualizar as palavras-chave, os sinônimos e os termos relacionados que foram empregados nesse protocolo:

Tabela 13 – Palavras-chave e termos relacionados para revisão da literatura

Palavras-chave	Sinônimos e/ou Termos relacionados
Machine learning	Supervised Learning
Fraud	Detection
Public sector	Public administration

^a Fonte: Elaboração própria.

Foram definidas duas expressões de busca. A primeira, a qual possuía o objetivo geral de identificar trabalhos relacionados à aprendizagem de máquina e que estivessem relacionados ao emprego dessas técnicas na detecção de fraude em contextos gerais, foi construída da seguinte forma:

(“Machine Learning” OR “Supervised Learning”) AND (“Fraud”) AND (“detect”)

Já com o propósito de buscar a aplicação dessas técnicas no contexto do setor público, as palavras-chave foram reordenadas conforme segunda expressão de pesquisa:

("machine learning"OR “Supervised Learning”) AND ("public sector"OR "public administration")

Destaca-se que a primeira expressão de busca foi realizada com enfoque nas áreas das ciências humanas, enquanto que a segunda teve o enfoque na produção nacional, dado o objetivo de se identificar aplicações relevantes no contexto brasileiro.

Após a implementação das pesquisas, foram obtidos 45 produções científicas, sobre os quais se realizou uma avaliação a fim de selecionar apenas aqueles que possuem relevância para a pesquisa conduzida.

Com esse objetivo, foram construídos os critérios de inclusão (CI) e critérios de exclusão (CE) para os artigos obtidos, os quais serviram como um filtro avançado.

Os critérios de inclusão utilizados foram os seguintes:

- **CI1:** serão incluídos apenas artigos completos relacionados ao tema, nos quais tragam no título ou resumo informações suficientes sobre o trabalho desenvolvido; e
- **CI2:** serão incluídos apenas artigos científicos que estejam em língua inglesa ou em língua portuguesa.

Como critérios para exclusão de trabalhos (CE), foram utilizados os listados a seguir:

- **CE1:** serão desconsiderados todos os documentos que não sejam artigos científicos;
- **CE2:** caso existam artigos repetidos, serão mantidos, apenas, os mais recentes;
- **CE3:** serão desconsiderados artigos em que o acesso completo não esteja disponível;
- **CE4:** serão desconsiderados artigos que foram retratados ou desconsiderados, se houver; e
- **CE5:** serão desconsiderados artigos em que os modelos computacionais não sejam aplicados no contexto de gestão, controle ou fiscalização de procedimentos, públicos ou privados.

Após o uso dos critérios de seleção, os artigos remanescentes, no caso 17, foram avaliados e as aplicações e proposições mais relevantes, integraram a seção de revisão da literatura.