

Controle Externo da Administração Pública: proposta de aplicação de mineração de textos para subsidiar a auditoria de conformidade em contratos

External Control of the Public Administration: proposed application of text mining to subsidize contract compliance audit

<https://doi.org/10.32586/rdca.v21i2.831>

Paulo Alcântara Saraiva Leão¹

RESUMO

O controle externo da administração pública, uma das etapas do ciclo da política pública, desempenha um papel relevante sendo um instrumento de acompanhamento da legalidade e desempenho da gestão pública. O controle orientado por dados pode contribuir para uma melhor gestão pública, exercida de forma legal e produtiva, e com políticas públicas efetivas e inovadoras, que proporcionem melhor qualidade de vida para a população. Os dados que são produzidos no desempenho de atividades públicas e privadas podem ser utilizados pelo controle da administração pública para aperfeiçoar sua missão e a qualidade das fiscalizações. A ciência de dados pode contribuir para o aperfeiçoamento do controle da administração pública. Os benefícios do seu uso no controle são significativos e permitem melhor identificação de fraudes e outras irregularidades ou riscos no uso dos recursos públicos, por meio da detecção de padrões e da revelação de *insights* nos dados, bem como a previsão de situações futuras indesejáveis. O objetivo deste trabalho é apresentar uma proposta de aplicação de mineração de textos, para subsidiar a auditoria de conformidade em contratos celebrados pela administração pública. A metodologia utilizada na pesquisa foi bibliográfica com o emprego do método indutivo para estudar

¹ Mestre em Ciências em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro (UFRJ), MBA em Gestão Empresarial pela Fundação Getúlio Vargas, especialista em Ciências de Dados pelo Centro Universitário 7 de setembro (UNI7) e em Políticas Públicas para Cidades Inteligentes pela Universidade de São Paulo (USP), bacharel em Computação pela Universidade Federal do Ceará (UFC). Atualmente é Gerente de Pesquisa, Inovação e Gestão da Informação do Instituto Plácido Castelo do Tribunal de Contas do Estado do Ceará. E-mail: paulo.alcantara65@gmail.com

aplicações e possibilidades de utilização da ciência de dados, em especial da mineração de textos, no escopo do controle.

Palavras-chave: ciência de dados; controle externo da administração pública; controle governamental; mineração de textos.

ABSTRACT

The external control of public administration, one of the stages of the public policy cycle, plays an important role as an instrument for monitoring the legality and performance of public management. Data-driven control can contribute to better public management, exercised in a legal and productive manner, and with effective and innovative public policies that provide a better quality of life for the population. The data that are produced in the performance of public and private activities can be used by public administration control to improve its mission and the quality of inspections. Data science can contribute to the improvement of public administration control. The benefits of its use in control are significant and allow better identification of fraud and other irregularities or risks in the use of public resources, through the detection of patterns and the revelation of insights in the data, as well as the prediction of undesirable future situations. The objective of this work is to present a proposal for the application of text mining, to subsidize the compliance audit in contracts entered into by the public administration. The methodology used in the research was bibliographic with the use of the inductive method to study applications and possibilities of using data science, especially text mining, in the scope of control.

Keywords: data science; external control of public administration; government control; text mining.

Avaliado pelo sistema
double blind review
(SEER/OJS – versão 3)



Data de submissão: 26-11-2022
Data de versão final: 10-04-2023
Data de aprovação: 16-04-2023
Data de publicação online: 19-06-2023

1 INTRODUÇÃO

Diante da velocidade das mudanças e aumento da complexidade dos problemas sociais contemporâneos, as abordagens tradicionais dos governos não são suficientes para solucioná-los e para atender as demandas existentes (CAVALCANTE, 2019). Vivemos uma era de alta complexidade, em que a realidade se apresenta com alto grau de volatilidade, incerteza e ambiguidade (MARQUESANI, 2020, p. 2). Nesse contexto, experimentamos a transformação digital, que impacta fortemente a forma com que nos relacionamos e produzimos e consumimos bens e serviços. Esse novo mundo traz imensos desafios para a sociedade e consequentemente para o Estado (BRASIL, 2020a).

Na administração pública, a ação de controlar diz respeito a verificar se as atividades realizadas estão em conformidade com sua finalidade pública e com a legislação aplicável ao setor público, bem como se seu desempenho é satisfatório, com resultados efetivos em prol da sociedade. Essa visão amplia a conceituação apresentada por Di Pietro (2020, p. 1659), quando define o controle da administração pública como sendo: “o poder de fiscalização e correção que sobre ela exercem os órgãos dos poderes Judiciário, Legislativo e Executivo, com o objetivo de garantir a conformidade de sua atuação com os princípios que são impostos pelo ordenamento jurídico”.

Nesse sentido, o controle contemporâneo busca garantir não somente a conformidade com a legislação, mas também a efetividade da gestão pública. Por meio de auditorias de desempenho (ou operacionais), os órgãos de controle avaliam os resultados alcançados pelas políticas públicas. É o que comenta Carvalho Júnior (2020): “o controle exercido pelos tribunais de contas não pode se resumir a meros relatórios contábeis e fiscalizações formais, devendo alcançar também a auditoria do desempenho das atividades de gestão”.

O tratamento científico dos dados pode contribuir para o aperfeiçoamento do controle da administração pública, uma vez que pode disponibilizar informações, de forma sistemática e baseada em critérios científicos, que poderão aprimorar o desempenho das ações de fiscalização, de forma mais produtiva, tempestiva, aprofundada e precisa (OLIVEIRA, 2017, p. 26).

O controle orientado por dados, com informações mais precisas, permite aos tribunais de contas trabalharem de forma mais estratégica e planejada, com priorização de auditorias e outras ações de controle em objetos de maior risco e com maior potencial de impacto (OLIVEIRA, 2017, p. 27).

A visão contemporânea do controle externo pode priorizar o controle preventivo, demonstrar proatividade, antecipar problemas, evitar ou mitigar desvios ou uso indevido de recursos públicos, quer por irregularidade ou por falta de efetividade. O uso intensivo de dados nas ações de controle pode contribuir para uma abordagem preventiva do controle, por meio de técnicas de análise preditiva (BALANIUK, 2017, p. 4).

O tratamento científico de dados pode dispor de uma extensa gama de informações e *insights*, que podem produzir o conhecimento sobre uma determinada realidade, referente aos relacionamentos, padrões, mudanças, tendências etc., obtida a partir do tratamento científico dos dados. Este tratamento aprimora as fiscalizações, com achados mais precisos, ações preventivas eficazes e tempestivas, responsabilizações mais objetivas e individualizadas, fiscalizações orientadas a risco, dentre outros benefícios (SOUZA, 2016; OLIVEIRA, 2017, p. 27). Um controle baseado em dados tende a ser mais efetivo, pois pode detectar, com mais precisão, irregularidades ou riscos na utilização dos recursos públicos, contribuindo para uma maior efetividade na definição e execução das políticas públicas e proporcionar melhoria na qualidade de vida das pessoas. A introdução do tratamento científico de dados na área do controle pode ser considerada uma inovação na gestão pública e no ciclo das políticas públicas.

Com base nas pesquisas realizadas, as atividades de controle da administração pública ainda são, em grande medida, realizadas com técnicas tradicionais, sem se beneficiar da totalidade dos dados e de recursos automatizados. A forma tradicional, considerada como “análogica” em relação ao modo digital de fiscalizar, utiliza pouco ferramentas de TI e ciência de dados, e trabalha com amostragem na seleção dos objetos auditados, ao invés de analisar todos os dados disponíveis, podendo limitar a eficácia das auditorias por reduzir a quantidade de objetos efetivamente verificados (OLIVEIRA, 2017, p. 26). A realização de auditorias da forma tradicional pode demandar muito esforço, a depender do escopo e quantidade de objetos a serem verificados. Embora a amostragem tenha embasamento científico para garantir uma boa representatividade do conjunto total de dados, é desejável, caso viável, a verificação de todo o conjunto, o que aumentaria a possibilidade de se obter mais achados nas auditorias.

O controle pode, portanto, beneficiar-se do emprego da ciência de dados para analisar conjuntos totais de objetos, aumentar o escopo das auditorias, automatizar verificações, e otimizar o tempo despendido pelos auditores. Nesse sentido, a utilização da mineração de texto, uma técnica de ciência de dados, pode beneficiar a realização de auditorias a partir de fontes em formato textual (dados não estruturados), por exemplo, os contratos (SOARES, 2020, p. 203). A partir dessa situação, foi construída a seguinte questão de pesquisa: como a ciência de dados, em especial a mineração de textos, pode contribuir para o aperfeiçoamento do controle de conformidade de contratos?

O objetivo geral dessa pesquisa é desenvolver uma proposta de aplicação da mineração de textos, disciplina da área de ciência de dados, para subsidiar auditorias de conformidade de contratos no Tribunal de Contas do Estado do Ceará, normatizadas por meio do Manual de Auditoria de Conformidade do Tribunal (Resolução Administrativa nº 5/2017) (CEARÁ, 2017). Essa pesquisa visa contribuir para o aperfeiçoamento do controle externo da administração pública.

A pesquisa constou de um levantamento bibliográfico de trabalhos científicos disponíveis na literatura sobre aplicações de ciência de dados, em especial da mineração de textos, no controle externo da administração pública nacional. O método utilizado foi o indutivo que permitiu desenvolver a questão de pesquisa proposta e a partir do estudo de experiências de aplicação da mineração de dados no controle em tribunais de contas. Conforme ensina Gil (2002), com relação aos procedimentos técnicos, a pesquisa bibliográfica é desenvolvida tendo como base materiais (artigos, livros etc.) já elaborados e publicados. Quanto à abordagem do problema, tratou-se de uma pesquisa qualitativa e de caráter exploratório.

2 REFERENCIAL TEÓRICO

2.1 O controle externo da administração pública

O controle da administração pública, previsto nos artigos 70 a 75 da Constituição Federal, tem a finalidade de garantir o bom uso dos recursos públicos, observando os princípios de legalidade, impessoalidade, moralidade, publicidade e eficiência na gestão pública. É efetivado por meio da fiscalização de natureza contábil, financeira, orçamentária, operacional e patrimonial da administração pública direta e indireta (BRASIL, 1988). Todos os gestores que administram recursos públicos devem prestar contas aos órgãos de controle e serem responsabilizados, em caso de desrespeito às normas vigentes. O controle da administração pública é dividido em externo e interno. Na definição de Guerra (2005, p. 108) o controle externo é “aquele desempenhado por órgão apartado do outro controlado, tendo por finalidade a efetivação de mecanismos, visando garantir a plena eficácia das ações de gestão dos interesses da sociedade, por órgão de fora de suas partes”.

No âmbito do controle externo, a fiscalização é uma atividade que tem como objetivo avaliar a gestão dos recursos públicos e consiste na

coleta de dados e informações, análise e formação de juízo de valor. É por meio dela que o controle é concretizado. Pode ser feita por iniciativa própria, por solicitação do Poder Legislativo ou para apuração de denúncias e representações. No Tribunal de Contas do Estado da União (TCU), e de forma geral nos demais tribunais, as fiscalizações podem ser realizadas por meio dos seguintes instrumentos: levantamento, auditoria, inspeção, acompanhamento e monitoramento (BRASIL, 2020b).

A auditoria, no contexto do controle da administração pública é, portanto, um tipo de fiscalização. Conforme as Normas de Auditoria Governamental (NAG), “auditoria é o exame independente, objetivo e sistemático de dada matéria, baseado em normas técnicas e profissionais, no qual se confronta uma condição com determinado critério com o fim de emitir uma opinião ou comentário” (CEARÁ, 2017).

Conforme Brasil (2001), a auditoria:

consiste no conjunto de técnicas que visa avaliar a gestão pública, pelos processos e resultados gerenciais, e a aplicação de recursos por entidade de direito público e privado, mediante a confrontação entre uma situação encontrada com um determinado critério técnico, operacional ou legal (BRASIL, 2001, p. 34).

Dentro do ambiente público existem três tipos de auditoria: financeira, operacional e de conformidade. A auditoria de conformidade é a “avaliação independente para determinar se um dado objeto está em conformidade com normas aplicáveis identificadas como critérios” (Instituto Rui Barbosa, 2017). Seu objetivo é avaliar se o que está sendo auditado, está de acordo com as normas que o regulamentam. Nas seções seguintes do presente trabalho, será explorada a aplicação da ciência de dados especificamente no âmbito da auditoria de conformidade.

O controle da administração pública avalia a qualidade das políticas públicas postas em prática pela administração pública. Segundo conceitua Ribeiro (2017), políticas públicas representam um conjunto de ações e

decisões governamentais, capazes de promover modificações na realidade social e atribuir direitos e responsabilidades em prol de determinado objetivo de interesse público.

As políticas públicas podem ser estruturadas em um ciclo com várias etapas de seu desenvolvimento. Agum (2015) sugere que este ciclo seja composto das seguintes etapas: problema, agenda, alternativas, tomada de decisão, implementação e avaliação. De forma geral, o controle externo da administração pública participa e contribui na etapa de avaliação das políticas, juntamente com outros atores (cidadãos e organizações), na realização de auditorias sobre os resultados alcançados mediante a execução pela administração pública. Ressaltando explicitamente o papel dos tribunais de contas no ciclo da política pública, Duarte (2013) introduz mais uma etapa após a de avaliação, a de fiscalização e controle da execução da política por meio da atuação da sociedade civil, dos Tribunais de Contas e do Ministério Público.

2.2 Ciência de dados

A ciência de dados (em inglês: *data science*) é um campo de estudo interdisciplinar que objetiva a extração de conhecimento, detecção de padrões e obtenção de *insights* (descobertas de informações não triviais), a partir de dados, para a tomada de decisão. Segundo Provost e Fawcett (2013, p. 5, tradução nossa), “a ciência de dados envolve princípios, processos e técnicas para compreender fenômenos através da análise automatizada de dados”².

O escopo da ciência de dados compreende a identificação, coleta, transformação, processamento de dados e apresentação de informações úteis ao negócio, que visa permitir a análise e compreensão de comportamentos, relacionamentos e características, tomada de decisão e a previsão de cenários futuros (SOARES, 2020, p. 205).

² No original: “*Data science involves principles, processes and techniques for understanding phenomena via the (automated) analysis of data.*”

Técnicas de ciência de dados podem ser aplicadas ao contexto do controle da administração pública, notadamente nas atividades de auditoria de conformidade, na busca por padrões, atipicidades, correlações e tendências, a partir dos dados relativos à utilização de recursos pela administração pública. O uso dessas técnicas pode contribuir com a prevenção e detecção de fraudes, detecção da acumulação ilegal de cargos públicos e pagamento de benefícios indevidos, detecção de cartéis e conluios, análise da eficiência de programas, avaliação de políticas públicas, combate à corrupção, auditoria em licitações e contratos e convênios, análise de jurisprudência, avaliação de despesas e receitas públicas dentre outras. Além disso, podem estimular uma maior governança e transparência por parte da gestão pública (SOARES, 2020, p. 197).

O emprego das técnicas de ciência de dados pode suportar o monitoramento da eficiência, eficácia e legalidade de despesas públicas, e fortalecer o controle sobre a área pública de forma geral, tanto por órgãos de controle quanto pela população quando exerce o controle social. Por meio delas, o auditor pode ter um conhecimento mais aprofundado das organizações fiscalizadas e, como consequência, aprimorar o processo de avaliação de riscos (SOARES, 2020, p. 205).

Com a ciência de dados, torna-se possível avaliar controles e riscos de maneira mais precisa, automatizar processos repetitivos e liberar os auditores da análise de processos manuais, tarefa que consome muito tempo sem, muitas vezes, agregar valor (SOARES, 2020, p. 205). Com a avaliação automatizada, pode-se ampliar não apenas a abrangência da auditoria, mas também a frequência das verificações. Isso permite viabilizar de forma prática o conceito da auditoria contínua (VASARHELYI, 1991), que busca auditar uma situação ou fato (objeto auditado) o mais próximo possível do momento de sua realização (em tempo real).

2.3 Mineração de dados

A mineração de dados é definida por Fayyad, Piatetsky-Shapiro e Smyth (1996, p. 39) como a aplicação de algoritmos específicos para a extração de padrões de dados. Para Han (2000, p. 7), *data mining* é o processo de descoberta de conhecimento útil em grandes volumes de dados armazenados em bancos de dados ou outros repositórios de dados. Os padrões, relações e associações encontradas em grandes volumes de dados podem revelar fatos importantes e tendências, para previsão de contextos e cenários, dentro de um processo de tomada de decisão.

Com a mineração de dados, é possível realizar, por exemplo, análises preditivas que permitem o processamento dos dados em busca de indícios, riscos e tendências. A análise preditiva permite que os auditores possam tomar ações preventivas, com vistas a evitar o dano por meio da detecção de fraudes, pagamentos indevidos e atividades criminosas. É um controle proativo que almeja a detecção de falhas e irregularidades antes da utilização do recurso público (BALANIUK, 2017, p. 295).

A mineração de dados pode contribuir para otimizar as ações de controle e possibilita a fiscalização sobre objetos com maior potencial de causar prejuízos ou irregularidades à administração pública (orientação a risco) (BALANIUK, 2010, p. 82).

2.4 Mineração de textos

A mineração de textos é um tipo de mineração de dados aplicada à busca de conhecimentos em textos, que são essencialmente dados não estruturados (MELO, 2016). Segundo Aranha (2006, p. 2) “mineração de textos é um conjunto de métodos usados para navegar, organizar, achar e descobrir informação em bases textuais”. Já Feldman (1997) define como sendo “uma extração não trivial de informações, não explícitas, de grandes bases textuais, previamente desconhecidas, e potencialmente úteis”.

O principal objetivo da mineração de texto é descobrir termos relevantes em conteúdos textuais, assim como, definir padrões e classificar o material por assunto ou tema a partir da frequência e temática dos termos encontrados dentro do domínio analisado (SERAPIÃO, 2010, p. 104).

O processo de mineração de texto compreende basicamente quatro etapas: coleta de documentos, pré-processamento, extração de conhecimento e avaliação e interpretação dos resultados (MARTINS, 2003, p. 2).

A coleta de documentos tem como objetivo recuperar conteúdo relevante ao contexto que se deseja analisar. Na etapa de pré-processamento se dá a preparação dos conteúdos coletados em linguagem natural, para serem processados pelos algoritmos de extração de conhecimento. Nessa terceira etapa, o objetivo é descobrir padrões desconhecidos presentes nos documentos. Na etapa final, o processo é avaliado e os padrões identificados são analisados por especialistas do domínio, que buscam a interpretação dos resultados obtidos (MARTINS, 2003, p. 2).

Grande parte das informações utilizadas por auditores está contida em textos, como: normas, atas, relatórios, atos, contratos, convênios, pareceres, denúncias, deliberações, publicações em diário oficial, prestações de contas, dentre outros. Ao lidar intensivamente com processos administrativos, o auditor trabalha fortemente com informações no formato textual, processo que demanda muito tempo e esforço para análise manual desse material (BALANIUK, 2017, p. 10). Esse trabalho processual pode ser otimizado por técnicas de mineração de textos. Ainda para Balaniuk (2017, p. 12), a utilização da mineração de texto nas atividades de controle tem potencial para agilizar o trabalho processual e, principalmente, para melhorar sua qualidade e assertividade.

Um exemplo clássico de aplicação de mineração de texto em auditoria de conformidade, a ferramenta Alice (Analizador de Licitações, Contratos e Editais) analisa informações textuais de editais de licitações e atas de registro de preços publicados em diários oficiais e outras fontes. Silva (2016) relata trabalhos de fiscalização do TCU que utilizam téc-

nicas de análises de dados e, dentre eles, descreve a Alice, uma solução tecnológica desenvolvida no Tribunal, para fornecer informações para as auditorias. A solução envia diariamente mensagens de alerta automáticas aos auditores indicando riscos e indícios de fraudes e irregularidades nas licitações e atas de registro de preços publicadas no dia anterior, a partir da análise textual de editais considerando tipologias, tais como, valores envolvidos e indícios de possíveis irregularidades. Costa (2020) explica que cada tipologia caracteriza um padrão que pode representar um indício de irregularidade ou risco na licitação, por exemplo, a exigência de uma certidão negativa de protesto, o que poderia acarretar uma restrição de competitividade durante a fase de habilitação de licitantes.

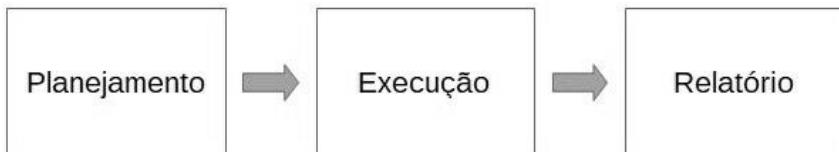
Amaral (2020) propõe um modelo para segmentação de itens adquiridos pelo Governo de Estado de Pernambuco, a partir de características similares extraídas de suas descrições encontradas em bases cadastrais, para disponibilizar informações úteis para ações de controle. A técnica utilizada para a extração das informações textuais foi a de alocação de tópicos latentes, que tem como objetivo agrupar documentos de acordo com seus temas, a partir da identificação de tópicos (grupos de palavras que sugerem um tema ou assunto em um documento).

Outra experiência do emprego da mineração de textos para apoiar auditorias é a pesquisa relatada por Melo (2016) com o objetivo de demonstrar a viabilidade da implementação de técnicas de mineração de dados para detecção de indícios de irregularidades em auditorias governamentais. Foram utilizados dados relativos a contratações públicas, com objetivo de identificar fraudes e risco de ocorrência das seguintes irregularidades: ausência de competição em licitações realizadas, fracionamento de despesas, cartel ou rodízio de empresas em licitações e sobre licitantes distintos que possuem sócios ou endereços em comum.

3 O PROCESSO DE AUDITORIA NO TRIBUNAL DE CONTAS DO ESTADO DO CEARÁ (TCE)

Um processo clássico de auditoria no setor público possui três fases consecutivas: planejamento, execução e relatório (CARVALHO NETO, 2011, p. 45). O processo de auditoria no TCE está definido por meio da Portaria nº 251/2021 (CEARÁ, 2021) e também compreende três fases: planejamento, execução e relatório. Na fase de planejamento são definidas as estratégias e a programação dos trabalhos. Na fase de execução o trabalho propriamente dito é realizado por meio da coleta de dados e exame de evidências para fundamentar o parecer. Na fase de produção do relatório, é formalizado o resultado da auditoria com a comunicação sobre o que foi examinado e encontrado (CEARÁ, 2021, p. 5), conforme Figura 1 abaixo.

Figura 1 – Fases de uma auditoria



Fonte: elaborada pelo autor (2022).

Periodicamente o TCE, a partir do planejamento estratégico institucional, define as diretrizes para as ações de controle, estabelece os temas de maior relevância a serem trabalhados e define os critérios para a elaboração do plano de fiscalização que inclui as auditorias e outros instrumentos de controle priorizados (CEARÁ, 2017, p. 21).

O plano é elaborado de uma matriz de riscos que sugere objetos que devem ser priorizados para serem auditados por apresentarem indícios ou alto risco de irregularidade. Além do plano de fiscalização, as auditorias podem ser autorizadas e iniciadas a partir de decisões colegiadas do Tribunal, sorteios aleatórios de órgãos e entidades públicas e municípios ou alguma situação relevante para o controle.

4 PROPOSTA METODOLÓGICA PARA UTILIZAÇÃO DE MINERAÇÃO DE TEXTOS PARA AUDITORIAS EM CONTRATOS NO TCE

Este trabalho propõe a utilização da mineração de texto, área da ciência de dados, no âmbito da Resolução Administrativa nº 5/2017, que visa aprimorar o processo de auditoria de contratos no TCE. A aplicação de mineração de texto em contratos irá disponibilizar informações (não triviais e desconhecidas) e *insights* para os auditores, bem como contribuirá para melhor elaboração do plano de fiscalização, quando as auditorias são definidas e priorizadas, e para o planejamento de uma auditoria em particular.

A mineração de texto pode ser empregada para extrair indícios de possíveis irregularidades ou não conformidades com padrões pré-definidos em contratos para subsidiar auditorias e outras ações de controle. Nas fases de planejamento e execução das auditorias, a aplicação da mineração de texto é útil para levantar possíveis indícios de irregularidades ou riscos, e indicar pontos que mereçam atenção ou aprofundamento na análise, coletar evidências e confirmar achados.

Conforme o processo de auditoria de conformidade do TCE, normatizado por meio da Resolução Administrativa nº 05/2017, os resultados da mineração de texto podem ser considerados fontes de informação que representam condições ou situações reais encontradas em objetos de auditoria (CEARÁ, 2017, p. 20). Tais fontes são insumos para as fases da auditoria propriamente dita.

No processo proposto, a totalidade dos contratos pode ser verificada e não apenas amostras da maneira que é realizado tradicionalmente e previsto nas normas internacionais ISSAI 400 – Princípios Fundamentais de Auditoria de Conformidade (*International Organization of Supreme Audit Institutions*, 2017) e ISSAI 1530 – Amostragem de Auditoria (*International Organization of Supreme Audit Institutions*). A introdução dessa verificação, facilitada pela mineração de textos, permite que toda a população de dados

de interesse das auditorias seja analisada de forma automatizada, o que cobre escopos maiores e pode elevar a produtividade da força de trabalho.

Os dados sobre contratos, em formato textual, são obtidos em publicações em diários oficiais (em conteúdos resumidos chamados de extratos) além de bases documentais. Essas bases possuem o conteúdo integral dos contratos, tais como o sistema de contrato do TCE e o Portal da Transparência do Estado. Nessas fontes podem ser encontrados aditivos de contratos e pagamentos realizados, que podem ser processados para busca de indícios. A mineração de texto pode ser utilizada em conjunto com cruzamentos de metadados e dados estruturados armazenados em bancos de dados hospedados no ambiente computacional do TCE, como por exemplo, nos sistemas internos de contratos. Essa técnica não precisa ser empregada de forma isolada, mas pode ser combinada com técnicas de mineração de dados, o que amplia o alcance das verificações e assim enriquece o controle.

O emprego da mineração de texto no âmbito do TCE pode seguir a metodologia abaixo proposta:

1. coleta de conteúdos textuais sobre a totalidade dos contratos da administração pública (e seus aditivos) a partir de dados extraídos de edições publicadas nos diários oficiais e em bases documentais de contratos disponíveis no TCE. Nesse passo, pode-se utilizar técnicas de raspagem (Saurkar; Pathare; Gode, 2018) de texto em documentos em páginas web;

2. limpeza e formatação de maneira padronizada dos textos coletados e armazenamento em forma de dados estruturados. Nesse passo o conteúdo extraído é preparado para ser processado por algoritmos no passo seguinte;

3. aplicação de algoritmos de busca de informações não triviais e insights nos dados, que sejam de interesse para a auditoria de contratos, confrontadas com critérios de auditoria (tipologias) que representem ilicitudes. Esse passo visa identificar indícios de possíveis irregularidades ou riscos;

4. avaliação e análise dos resultados obtidos pela etapa anterior para verificar se os indícios são fortes e justificam a realização de auditorias;

5. geração de relatório com recomendações, sugestões e pontos de atenção. O relatório será disponibilizado aos auditores ou à equipe de elaboração do Plano de Fiscalização, a depender da fase da auditoria onde a mineração de texto foi inserida.

Dentre as principais técnicas de mineração de texto, propomos a utilização da extração de informação que tem como objetivo a extração de informações específicas a partir de documentos de texto (MACHADO, 2010, p. 14).

Como exemplos de possíveis tipologias a serem aplicadas ao conteúdo textual de contratos temos:

- itens raros ou não adequados adquiridos pelo órgão contratante;
- itens não compatíveis com o objeto do contrato;
- contrato assinado com empresa que possui como sócio algum servidor do órgão contratante (BALANIUK, 2010, p. 81);
- contrato celebrado com empresa proibida de contratação com a administração pública ou empresa fantasma (COSTA, 2020);
- objeto do contrato não compatível com o objeto social da empresa contratada;
- capital social da empresa incompatível com o valor do contrato;
- empresa contratada sem funcionário na Relação Anual de Informações Sociais (RAIS).

Esses critérios são exemplos de possibilidade de extração de informações dos textos alertando quanto às possíveis irregularidades, recurso que irá prover subsídios para auditorias. As tipologias podem ser adequadas aos tipos de contratação e às áreas relativas ao objeto contratado.

Outra possibilidade é verificar nos textos a existência ou falta de cláusulas (e o que estabelecem) que precisam constar dos contratos, para garantir sua legalidade e identificar se as condições acordadas não são prejudiciais à administração pública.

Consideramos uma inovação no TCE introduzir uma atividade automatizada no processo de auditoria, baseada em ciência de dados, com mineração de texto. Essa aplicação tem o potencial para aumentar a produtividade do processo, aumentar o escopo de auditoria e melhorar a quantidade e qualidade de achados, ao mesmo tempo em que permite que os auditores possam se dedicar a tarefas de análises mais complexas que precisam da análise humana.

4.1 Potenciais benefícios mapeados com a implantação do processo de mineração de texto

A introdução da mineração de textos no processo de auditoria de conformidade do TCE, com a verificação de tipologias diretamente nos contratos da administração pública, em conteúdo textual e integral, vai complementar a verificação que já é feita a partir de dados estruturados sobre contratos armazenados em bases de dados. Algumas informações importantes para a auditoria não estão disponíveis em dados cadastrais sobre contratos. Dessa forma, a descoberta de informações úteis ao controle, diretamente nos textos dos contratos, agrupa valor à auditoria desses instrumentos.

Conforme abordado na literatura pesquisada e apresentado na proposta de aplicação prática no TCE, a mineração de textos aplicada às auditorias de contratos pode contribuir para melhor identificação de possíveis indícios de irregularidades e riscos detectados em documentos e bases de dados textuais. Essa aplicação tem potencial para aperfeiçoar a priorização de auditorias e contribuir com informações durante as fases de planejamento e execução das auditorias.

A auditoria de contratos tem uma grande importância no controle da administração pública, pois diversas e numerosas irregularidades no uso de recursos públicos se dão justamente na execução de contratos, quando valores são transferidos a agentes privados.

A automatização permite que a análise de contratos seja realizada com mais frequência, possibilitando a detecção de possíveis problemas de forma mais tempestiva, pois a rotina de verificação pode ser executada, por exemplo, diariamente.

Os principais benefícios potenciais mapeados com a implantação do processo de utilização de mineração de texto proposto neste trabalho, como apoio ao planejamento e execução de auditorias de conformidade, foram:

- orientação ao planejamento de auditorias;
- aprimoramento da avaliação baseada em riscos no planejamento de auditoria, por meio da identificação de riscos relevantes, para priorização dos objetos a serem auditados;
- melhoria na identificação de padrões, classes, correlações, anomalias, exceções, *outliers* e tendências nos dados de contratos analisados;
- identificação de *insights* a partir dos dados coletados, o que pode revelar informações desconhecidas ou não identificáveis por meio de análises triviais;
- possibilidade da análise de conjuntos completos de dados ao invés de amostras, como é feito nos métodos tradicionais de auditoria;
- melhoria da assertividade das auditorias, o que as tornam mais precisas, baseadas em evidências científicas;
- possibilidade de viabilizar a sustação de despesas indevidas antes que ocorram, devido ao controle realizado antecipadamente, em tempo de publicação nos diários oficiais, e *pari passu*.

5 CONSIDERAÇÕES FINAIS

A presente pesquisa apresentou o potencial da utilização da ciência de dados, em particular da mineração de textos, no campo do controle externo da administração pública. Tanto no combate à corrupção quanto

na detecção de ineficiências da máquina pública, a ciência de dados se apresenta como um importante instrumento de suporte ao controle governamental com consequências benéficas para a gestão pública e para a sociedade. Dessa forma, um controle que utiliza ciência de dados pode contribuir positivamente com a melhoria da prestação dos serviços públicos e da execução das políticas públicas, além de fomentar a fiscalização que a sociedade pode exercer sobre os gestores públicos.

Foi desenvolvida uma proposta metodológica de aplicação da mineração de textos para subsidiar as auditorias de conformidade em contratos no TCE. A introdução da mineração de textos no processo formal de auditoria de conformidade pode produzir informações e *insights* sobre indícios de irregularidades em contratos que servirão de subsídios para o planejamento e execução de fiscalizações. A mineração identificará contratos com alto risco de irregularidades, a partir de tipologias definidas pelos auditores, e que merecem ser investigados mais profundamente.

Com a metodologia proposta, a totalidade dos contratos poderá ser verificada e essa verificação poderá ser feita de forma mais tempestiva. O aumento da precisão das auditorias também é um benefício esperado.

A partir do estudo realizado, concluímos que o emprego da mineração de textos pode trazer benefícios para auditorias de conformidade, ampliar o escopo das fiscalizações, melhorar a tempestividade das auditorias, elevar a assertividade e trazer maior confiabilidade aos resultados alcançados, além de aumentar a produtividade das atividades de controle.

O uso da mineração de textos no TCE é um campo de aplicação pouco explorado, tanto na área de auditoria de contratos quanto em outras áreas do controle, o que abre oportunidade para estudos e pesquisas. A partir da presente investigação, trabalhos futuros podem ser desenvolvidos como complemento ao estudo realizado, tais como a aplicação da metodologia proposta neste trabalho, de forma a validar sua contribuição efetiva, e o estudo sobre a aplicação de mineração de textos em convênios firmados entre órgãos e entidades públicas, e em decisões do Tribunal, o que pode permitir

a construção e manutenção de um banco de jurisprudências para auxiliar no processo de tomada de decisão no contexto do controle.

REFERÊNCIAS

AGUM, R.; RISCADO, P.; MENEZES, M. Políticas Públicas: Conceitos e Análise em Revisão. **Revista Agenda Política**, [S. l.], v. 3, n. 2, p. 12-42, jul./dez., 2015. Disponível em: <https://www.agendapolitica.ufscar.br/index.php/agendapolitica/article/view/67>. Acesso em: 27 abr. 2022.

AMARAL, J. A. da S.; RODRIGUES, J. B. Alocação de Tópicos Latentes - Um Modelo para Segmentação de Dados de Auditoria do Governo de PE. **Revista de Engenharia e Pesquisa Aplicada**, Recife, v. 5, n. 1, p. 40-49. 2020. DOI: 1025286/repa.v5i1.1179. Disponível em: <http://revistas.poli.br/index.php/repa/article/view/1179/573>. Acesso em: 27 abr. 2022.

ARANHA, C.; Passos, E. Tecnologia de Mineração de Textos (Artigo tutorial). **Revista Eletrônica de Sistemas de Informação (RESI)**, Curitiba, v. 5, n. 2, p. 1-8, 2006. Disponível em: <http://www.periodicosibepes.org.br/index.php/reinfo/article/view/171/66>. Acesso em: 27 abr. 2022.

BALANIUK, R. A Mineração de Dados como apoio ao Controle Externo. **Revista do Tribunal de Contas da União**, Brasília, ano 42, n. 117, p. 77-84, jan./abr., 2010. Disponível em: <https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/282>. Acesso em: 27 abr. 2022.

BALANIUK, R. Novas tecnologias aplicadas ao controle externo da administração pública. In: OLIVEIRA, A. C. de. **O controle da administração pública na era digital**. 2. ed., Belo Horizonte: Fórum, 2017. p. 291-317. Disponível em: <https://www.forumconhecimento.com.br/livro/1471>. Acesso em: 27 abr. 2022.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal: Centro Gráfico, 1988.

BRASIL. Ministério da Fazenda. Secretaria Federal de Controle Interno. **Manual do Sistema de Controle Interno do Poder Executivo Federal. Anexo à Instrução Normativa nº 1, de 06/04/2001**. Brasília, 2001. Disponível em: <http://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?fileId=8A8182A14D8D5AA6014D8D815F07258B>. Acesso em: 22 mar. 2020.

BRASIL. Ministério do Desenvolvimento Regional. **Carta Brasileira para Cidades Inteligentes**. Brasília, 2020a. Disponível em: https://www.gov.br/mdr/pt-br/assuntos/desenvolvimento-regional/projeto-andus/Carta_Bras_Cidades_Inteligentes_Final.pdf. Acesso em: 24 maio 2022.

BRASIL. Tribunal de Contas da União. **Conheça o TCU**. Brasília: TCU, 2020b. Disponível em: <https://portal.tcu.gov.br/institucional/conheca-o-tcu/funcionamento/>. Acesso em: 21 mar. 2020.

CEARÁ (Estado). Tribunal de Contas do Estado do Ceará. **Resolução nº 5, de 13 de junho de 2017**. Dispõe sobre a aprovação do Manual de Auditoria de Conformidade do Tribunal de Contas do Estado do Ceará. Fortaleza, 2017. Disponível em: <https://www.tce.ce.gov.br/exercicios-anteriores/resolucoes-administrativas/2017/send/247-resolucoes-administrativas-2017/3530-resolucao-administrativa-5-2017>. Acesso em: 1 maio 2022.

CEARÁ (Estado). **Resolução nº 251, de 08 de junho de 2021**. Aprova o Referencial aplicável às fases e aos documentos da instrução processual no âmbito da Secretaria de Controle Externo (SECEX) do Tribunal de Contas do Estado do Ceará (TCE/CE). Fortaleza, 2021. Disponível em: https://www.tce.ce.gov.br/downloads/RH/Portarias/2021/Portaria2512021_.pdf. Acesso em: 1 maio 2022.

CARVALHO JÚNIOR, N. E. de; TAVARES NETO, J. Q. Controle externo de políticas públicas: dificuldades e propostas de aperfeiçoamento da atividade dos Tribunais de Contas. *In: Revista Interesse Público – IP*, Belo Horizonte, ano 22, n. 120, p. 243-262, mar./abr. 2020. Disponível em: <https://www.forumconhecimento.com.br/periodico/172/41921/91570>. Acesso em: 1 maio 2022.

CARVALHO NETO, A. A. de (coord.). **Auditoria Governamental**. Brasília: TCU, 2011.

CAVALCANTE, P.; MENDONÇA, L.; BRANDALISE, I. Políticas públicas e Design Thinking: Interações para enfrentar desafios contemporâneos. *In: CAVALCANTE, Pedro (org.). Inovação e políticas públicas: superando o mito da ideia*. Brasília: Instituto de Pesquisa Econômica Aplicada (Ipea), 2019. p. 29-52. Disponível em: <https://repositorio.ipea.gov.br/bitstream/11058/9383/1/Pol%C3%ADticas%20p%C3%BAblicas%20e%20design.pdf>. Acesso em: 26 mar. 2023.

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. **Controle Externo**: Revista do Tribunal de Contas do Estado de Goiás, Belo Horizonte, ano 2, n. 3, p. 11-34, jan./jun., 2020. Disponível em: <https://revcontext.tce.go.gov.br/index.php/context/article/view/59>. Acesso em: 1 maio 2022.

CUNHA, I. B. de A.; BARACHO, R. M. A.; Dados Abertos e suas aplicações em Cidades Inteligentes. **Liinc em Revista**, [S. l.], v. 15, n. 2, 2019. DOI: 10.18617/liinc.v15i2.4767. Disponível em: <http://revista.ibict.br/liinc/article/view/4767>. Acesso em: 28 jun. 2021.

DI PIETRO, M. S. Z. **Direito Administrativo**. 33. ed. Rio de Janeiro: Editora Forense, 2020.

DUARTE, C. S. **O ciclo das políticas públicas.** In: SMANIO, G. P.; BER-TOLIN, P. T. M. *O Direito e as Políticas Públicas no Brasil*. São Paulo: Atlas, 2013. p. 16-43.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, p. 37, 15 mar. 1996. Disponível em: <https://doi.org/10.1609/aimag.v17i3.1230>. Acesso em: 17 maio 2022.

FELDMAN, R.; HIRSH, H. “Exploiting Background information in Knowledge discovery from text”. **Journal of Intelligent Information System**, v. 9, n. 1, 1997.

GIL, A. C. **Como elaborar projetos de pesquisa.** 4. ed. São Paulo: Atlas, 2002.

GUERRA, E. M. **Os controles externo e interno da administração pública.** 2. ed. Belo Horizonte: Editora Fórum, 2005.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques.** Simon Fraser University. Manuscript. 2000.

INSTITUTO Rui Barbosa **Normas Brasileiras de Auditoria do Setor Público (NBASP).** Nível dois: princípios fundamentais de auditoria do setor público. Belo Horizonte, 2017. 190p.

INTOSAI. International Organization of Supreme Audit Institutions. **Principios fundamentais de auditoria de conformidade:** ISSAI 400. [Brasília]: Tribunal de Contas da União - TCU, 2017. 23p.

INTOSAI. International Organization of Supreme Audit Institutions. Muestreo de auditoría: ISSAI 1530. **Directriz de auditoría financiera:** ISSAI 1000-2999. Estocolmo: INTOSAI. p. 675-702.

MACHADO, A. P. *et al.* **Mineração de Texto em Redes Sociais Aplicada à Educação a Distância**, 2010. Disponível em: <https://silo.tips/download/mineraao-de-texto-em-redes-sociais-aplicada-a-educaao-a-distancia>. Acesso em: 26 mar. 2023.

MARQUESANI, C. **Estilos de Liderança e Transformação Digital**: uma revisão de literatura. In: Congresso Transformação Digital. 2020. Disponível em <http://bibliotecadigital.fgv.br/ocs/index.php/ctd/ctd2020/paper/viewFile/7612/2312>. Acesso em: 28 jun. 2021.

MARTINS, C. A. *et al.* Uma experiência em mineração de textos utilizando clustering probabilístico clustering hierárquico. n. 205. **Relatórios Técnicos do ICMC**. Instituto de Ciências Matemáticas e de Computação. ISSN: 0103-2569. São Carlos, 2003.

MELO, B. M. S. S. (coord.). Descoberta de Conhecimento em Bases de Dados, com base em técnicas de Análise Preditiva, visando ao auxílio e ao aprimoramento da atividade de controle externo desempenhada pelo TCE-RJ. **Relatório final de projeto de pesquisa**. Programa de Pesquisa. Rio de Janeiro:ECG-TCE/RJ, 2016.

OLIVEIRA, A. C. de (coord.). **O controle da administração pública na era digital**. 2. ed. Belo Horizonte: Fórum, 2017. 460p.

PROVOST, F.; FAWCETT, T. Data Science and its relationship to big data and data-driven decision making. **Big Data**, v. 1, n. 1, p. 51-59, mar., 2013. Disponível em: <https://www.liebertpub.com/doi/10.1089/big.2013.1508>. Acesso em: 20 maio 2022.

RIBEIRO, A. C. R. N. **O Controle Externo de Políticas Públicas realizadas pelo TCU**: estudo de casos. 130f. Dissertação (Programa de Pós-

-Graduação Stricto Sensu em Direito e Políticas Públicas) – Centro Universitário de Brasília, UNICEUB, Brasília-DF, 2017.

SAURKAR, A. V.; PATHARE, K. G.; GODE, S. A. An overview on web scraping techniques and tools. **International Journal on Future Revolution in Computer Science & Communication Engineering**, v. 4, n. 4, p. 363-367, 2018.

SERAPIÃO, P. R. B.; SUZUKI, K. M.; MARQUES, P. M. de A. Uso de mineração de texto como ferramenta de avaliação da qualidade informatacional em laudos eletrônicos de mamografia. **Radiol Bras.** v. 43, n. 2, p.103-107, mar./abr., 2010. Disponível em: http://www.rb.org.br/edicao_atual.asp?ed=2. Acesso em: 20 jun. 2022.

SILVA, W. V. Os pilares da estratégia de análise de dados e consumo de informações no TCU. **Revista do Tribunal de Contas da União**, Brasília, ano 48, n. 137, p. 13-16, set./dez., 2016. Disponível em: <https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/1387>. Acesso em: 20 jun. 2022.

SOARES, G. F. Ciência de dados aplicada à auditoria interna. **Revista da CGU**, Brasília, v. 12, n. 22, jul./dez., 2020. Disponível em: https://revista.cgu.gov.br/Revista_da_CGU/article/view/195. Acesso em: 20 jun. 2022.

SOUZA, R. M. de. Inteligência computacional aplicada ao controle externo: classificação de padrões utilizando redes neurais artificiais. **Revista do Tribunal de Contas da União**, Brasília, ano 48, n. 135, p. 34-43, jan./abr., 2016. Disponível em: <https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/1323>. Acesso em: 20 jun. 2022.

VASARHELYI, M. A.; HALPER, F. B. The continuous audit of online systems. **Auditing a Journal of Practice and Theory**, v. 10, n. 1, jan., 1991.