UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Ambiente Inteligente de Análise de Riscos de Fraudes em Contratos Públicos

Dissertação de Mestrado

Luan Bruno Barbosa de Souza Costa

São Cristóvão – Sergipe
2023

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Luan Bruno Barbosa de Souza Costa

Ambiente Inteligente de Análise de Riscos de Fraudes em Contratos Públicos

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Sergipe como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Methanias Colaço Rodrigues Júnior

São Cristóvão – Sergipe
2023

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL
UNIVERSIDADE FEDERAL DE SERGIPE

Costa, Luan Bruno Barbosa de Souza
C837a Ambiente Inteligente de Análise de Riscos de Fraudes em Contratos Públicos / Luan Bruno Barbosa de Souza Costa ; orientador Methanias Colaço Rodrigues Júnior. - São Cristóvão, 2023.
61 f.; il.

Dissertação (mestrado em Ciência da Computação) – Universidade Federal de Sergipe, 2023.

1. Inteligência artificial. 2. Crimes contra a administração pública. I. Rodrigues Júnior, Methanias Colaço orient. II. Título.

CDU 004.8

UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Ata da Sessão Solene de Defesa da Dissertação do Curso de Mestrado em Ciência da Computação-UFS.
Candidato: Luan Bruno Barbosa de Souza Costa

Em 19 dias do mês de dezembro do ano de dois mil e vinte três, com início às 14h, realizou-se na Sala de Seminários do PROCC da Universidade Federal de Sergipe, na Cidade Universitária Prof. José Aloísio de Campos, a Sessão Pública de Defesa de Dissertação de Mestrado do candidato **Luan Bruno Barbosa de Souza Costa**, que desenvolveu o trabalho intitulado: "***Criação e Avaliação de um Ambiente de Análise de Riscos de Fraudes em Contratos Públicos***", sob a orientação do Prof. Dr. Methanias Colaço Rodrigues Júnior. A Sessão foi presidida pelo Prof. Dr. Daniel Oliveira Dantas (PROCC/UFS), que após a apresentação da dissertação passou a palavra aos outros membros da Banca Examinadora, Prof. Dr. Elyson Adan Nunes Carvalho (UFS) e, em seguida, o Prof. Dr. Leonardo Nogueira Matos (Procc/UFS). Após as discussões, a Banca Examinadora reuniu-se e considerou o mestrando (a) APROVADO. Atendidas as exigências da Instrução Normativa 05/2019/PROCC, do Regimento Interno do PROCC (Resolução 67/2014/CONEPE), e da Resolução no 04/2021/CONEPE que regulamentam a Apresentação e Defesa de Dissertação, e nada mais havendo a tratar, a Banca Examinadora elaborou esta Ata que será assinada pelos seus membros e pelo mestrando.

Cidade Universitária "*Prof. José Aloísio de Campos",* 19 de dezembro de 2023.

gov.br Documento assinado digitalmente
METHANIAS COLACO RODRIGUES JUNIOR
Data: 26/12/2023 20:29:27-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. METHANIAS COLACO RODRIGUES JUNIOR
(PROCC/UFS)
Presidente

gov.br Documento assinado digitalmente
DANIEL OLIVEIRA DANTAS
Data: 19/12/2023 22:37:35-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Daniel Oliveira Dantas
(PROCC/UFS)
Examinador Interno

gov.br Documento assinado digitalmente
JOSE AELIO DE OLIVEIRA JUNIOR
Data: 19/12/2023 23:08:57-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. José Aelio de Oliveira Júnior
(UFS)
Examinador Externo

LICURGO JOSEPH MOURAO DE OLIVEIRA:71146555415
Assinado de forma digital por LICURGO JOSEPH MOURAO DE OLIVEIRA:71146555415
Dados: 2023.12.20 10:42:14 -03'00'

Prof. Dr. Licurgo Joseph Mourão de Oliveira
(PUC - Minas)
Examinador Externo

gov.br Documento assinado digitalmente
LUAN BRUNO BARBOSA DE SOUZA COSTA
Data: 26/12/2023 20:37:59-0300
Verifique em https://validar.iti.gov.br

LUAN BRUNO BARBOSA DE SOUZA COSTA
Candidato

AGRADECIMENTOS

A Deus, que me deu a vida, a inteligência e a força para realizar este trabalho.

A meus pais, Pedro e Terezinha, e irmãos, Ítalo e Nathan, que sempre me apoiaram e incentivaram em minha jornada acadêmica.

A minha esposa, Daiana, que me apoiou e compreendeu minhas ausências durante o período de pesquisa.

Aos meus professores, que me transmitiram conhecimento e me ajudaram a crescer como profissional e como pessoa.

Aos meus colegas de trabalho do MPSE, que me ajudaram a coletar dados e a realizar a pesquisa.

RESUMO

Contexto: O gerenciamento de recursos públicos está sujeito a atos ilegais, que privam a sociedade de usufruir desses recursos em sua integralidade. A identificação de tais atos, por sua vez, depende da análise de uma grande quantidade de dados. **Objetivo**: O objetivo do presente trabalho é a criação e avaliação de uma ferramenta capaz de reduzir o escopo de busca por irregularidades, por meio de sinalizações de possíveis fraudes (*redflags*) em contratos firmados pelo estado, no contexto dos servidores do Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) do Ministério Público do Estado de Sergipe (MPSE). **Metodologia**: Utilizando a metodologia de pesquisa-ação, o que incluiu a realização de um mapeamento sistemático, uma lista de tipologias de risco de fraude foi identificada e automatizada na aplicação Galactus Snitch, a qual foi avaliada por meio de uma pesquisa qualitativa. **Resultados**: O mapeamento sistemático realizado identificou 19 trabalhos segundo os critérios informados, no qual cerca de 80% destes buscavam indicar algum tipo de fraude em processos licitatórios, obtendo acurácias entre 72% e 99%, com diferentes técnicas para abordar o problema. A partir de então, o Galactus Snitch foi desenvolvido e o survey aplicado revelou uma alta aceitabilidade da ferramenta, para a qual 66.7% dos servidores afirmaram concordar totalmente que o sistema é capaz de reduzir em mais de 90% o escopo de busca por irregularidades, ao apontar possíveis esquemas fraudulentos. Além disso, 91.7% dos servidores afirmaram também concordar totalmente que as tipologias identificadas são de grande relevância para que o sistema atinja o objetivo esperado. **Conclusões**: Utilizando análise de dados de diferentes fontes de informação, o trabalho apresenta uma importante contribuição para o processo de identificação de esquemas fraudulentos em contratos públicos, à medida que, segundo os entrevistados, é capaz de reduzir o escopo de busca por fraudes.

Palavras-chave: *Data Science*, *Inteligência Artificial*, Corrupção, Tipologia Criminal, *Redflag*.

ABSTRACT

Context: The management of public resources is subject to illegal acts, which deprive society of the full use of these resources. The identification of such acts, in turn, depends on the analysis of a large amount of data. **Objective**: The objective of this work is the creation and evaluation of a tool capable of reducing the scope of search for irregularities, by means of signs of possible fraud (redflags) in contracts signed by the state, in the context of the servers of the Special Action Group to Combat Organized Crime (GAECO) of the Prosecutor of the State of Sergipe (MPSE). **Methodology**: Using the research-action methodology, which included the conduct of a systematic mapping, a list of fraud risk types was identified and automated in the Galactus Snitch application, which was evaluated through a qualitative survey. **Results**: The systematic mapping carried out identified 19 works according to the criteria informed, in which about 80% of these sought to indicate some type of fraud in bidding processes, obtaining accuracy between 72% and 99%, with different techniques to address the problem. Since then, the Galactus Snitch was developed and the survey applied revealed a high acceptability of the tool, for which 66.7% of the servers stated fully agree that the system is able to reduce by more than 90% the scope of search for irregularities, by pointing out possible fraudulent schemes. In addition, 91.7% of servers also stated that they fully agree that the types identified are of great relevance for the system to the expected goal. **Conclusions**: Using data analysis from different sources of information, the work presents an important contribution to the process of identifying fraudulent schemes in public procurement, as, according to the respondents, it is able to reduce the scope of the search for fraud.

Keywords: Data Science, Artificial Intelligence, Corruption, Criminal Typology, Redflag.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AS/COA	*Americas Society/Council of the Americas*
BI	*Business Intelligence*
CADE	Conselho Administrativo de Defesa Econômica
CCC	Índice de Capacidade de Combate à Corrupção
EQM	Erro Quadrático Médio
GAECO	Grupo de Atuação Especial de Combate ao Crime Organizado
FIESP	Federação das Indústrias de São Paulo
IPC	Índice de Percepção de Corrupção
ISP	Inteligência de Segurança Pública
MPSE	Ministério Público do Estado de Sergipe
MRQ	*Mapping Research Question*
PICO	População, Intervenção, Comparação e *Outcomes* (Resultado)
RQ	*Research Question*
SQ	*Survey Question*
SQL	*Structured Query Language*
SVM	*Support Vector Machine*
TCE-SE	Tribunal de Contas do Estado de Sergipe
UFS	Universidade Federal de Sergipe

SUMÁRIO

1.0 INTRODUÇÃO

Neste capítulo, será realizada uma breve contextualização relacionada ao tema da pesquisa, motivação, problemática, questões, objetivos e suposição que se pretende evidenciar.

1.1 CONTEXTUALIZAÇÃO

A gestão de recursos públicos em muitos países, assim como no Brasil, infelizmente está sujeita a atos ilícitos, que visam a utilização ou subtração dos mesmos recursos em benefício próprio. Dentre os crimes contra a administração pública mais comuns, conforme a legislação brasileira, estão a Corrupção, o Peculato, a Prevaricação e a Concussão. No contexto brasileiro, um estudo realizado pelo Departamento de Competitividade e Tecnologia (Decomtec) da Fiesp (Federação das Indústrias de São Paulo) revelou que os prejuízos econômicos e sociais causados pela corrupção no país chegam a R$69 bilhões de reais por ano (FIESP, 2010). Ao mesmo tempo, o Índice de Capacidade de Combate à Corrupção (CCC), que é elaborado pela entidade empresarial americana *Americas Society/Council of the Americas* (AS/COA) e pela consultoria britânica *Control Risks*, indica que, desde 2019, o Brasil vem caindo no ranking que mede a capacidade de cada nação de combater a corrupção (Simon e Aalbers, 2019). Além disso, o Brasil ocupa a 96ª posição no Índice de Percepção da Corrupção (IPC), organizado pela Transparência Internacional, e que ordena os países do mundo de acordo com o grau em que a corrupção é percebida a existir entre os funcionários públicos e políticos, num total de 180 países (Corruption Perceptions Index, 2020).

Essa dificuldade em combater os crimes contra a administração pública passa pela dificuldade de analisar um grande volume de dados referentes à movimentação de ativos públicos, muitas vezes dispersos em diferentes bases de dados. Em decorrência disso, boa parte dos processos investigativos sobre danos ao erário são originados em denúncias feitas pelos próprios cidadãos (Velasco et al, 2021). Entretanto, apesar da dificuldade imposta pelo grande volume de informações, é justamente o fato de boa parte dos serviços governamentais estarem armazenados (e até certo ponto disponíveis) no formato digital que torna possível sua análise por meio do uso de *Data Science* e *Data Analytics*.

Diante desse cenário, faz-se necessário o uso e aprimoramento de técnicas e ferramentas que visem detectar, identificar ou predizer a existência de potenciais crimes

contra a administração pública. Em muitos casos, essas deduções só podem ser extraídas a partir da análise unificada de bases de dados distintas. A coleta de informações em bases de dados heterogêneas, a fim de auxiliar no processo de tomada de decisão, já é amplamente utilizada no setor privado, no âmbito mundial. Por exemplo, na formação do *Credit Score* - um índice que determina o quanto é seguro fornecer crédito a um determinado consumidor (Investopedia, 2021).

1.2 Problemática e Suposição

Diante deste cenário, onde as informações necessárias ao início de um processo de investigação de fraudes contra a administração pública estão ocultas e dispersas em diferentes bases de dados, o problema maior em questão é averiguar a possível existência de uma ferramenta que seja capaz de reduzir o escopo de busca por esses atos ilícitos no contexto dos órgãos de investigação. Uma vez desenvolvida a ferramenta, é necessário avaliá-la nesse mesmo contexto, a fim de verificar sua capacidade de aplicação prática. A partir desta indagação, as seguintes questões de pesquisa (***R**esearch **Q**uestions*) são colocadas em discussão:

a) RQ1 – Como são analisados grandes volumes de dados em busca de fraudes contra a administração pública?;
b) RQ2 – Quais os principais indicativos da possível existência de fraudes contra a administração pública?;
c) RQ3 – É possível que um software possa reduzir em mais de 90% o escopo de busca de fraudes contra a administração pública ao analisar um grande volume de dados?.

Ato contínuo, diante das questões levantadas, uma suposição passível de investigação deve ser elaborada. A suposição em questão é: Um ambiente que integre diferentes bases de dados e identifique possíveis atos ilícitos é capaz de reduzir em mais de 90% o escopo de busca por fraudes contra a administração pública, a ponto de tornar o órgão de investigação capaz de iniciar procedimentos mais assertivos.

1.3 Objetivo Geral

Este trabalho teve como objetivo geral o aperfeiçoamento e avaliação qualitativa de um ambiente que integre diferentes bases de dados e sinalize possíveis fraudes contra a administração pública, a fim de dar suporte aos órgãos de investigação na busca e prevenção ativas de crimes.

1.4 Objetivos Específicos

Para possibilitar a realização do objetivo geral, podemos enumerar os seguintes objetivos específicos:

a) OE1 - Execução de um Mapeamento Sistemático do estado da arte de ferramentas e técnicas que busquem a identificação ou sinalização da existência de fraudes no contexto da administração pública;
b) OE2 - Identificação de potenciais tipologias criminais que sinalizem a possível existência de esquemas fraudulentos em contratos públicos;
c) OE3 - Construção e avaliação de um sistema de análise, fruto de pesquisa-ação avaliativa, que integre diferentes bases de dados e sinalize possíveis contratos com fraude através do uso de *redflags*.

1.5 Metodologia

Considerando o ponto de vista exploratório, inicialmente, foi realizado um mapeamento sistemático da literatura, publicado em (Souza et al, 2023), com o objetivo de identificar lacunas e tendências acerca da identificação automatizada de crimes contra a administração pública, bem como de levantar requisitos e diagnosticar a utilização dessas ferramentas.

A fim de complementar o conhecimento necessário sobre tipologias de risco de fraude comuns a serem identificadas pela ferramenta e potencializar os resultados, a implementação do trabalho teve também como insumo o *Guia de Combate a Cartéis em Licitação*, publicado pelo Conselho Administrativo de Defesa Econômica (CADE) (CADE, 2019).

Do ponto de vista da pesquisa aplicada, e como forma de avaliar e melhorar a plataforma, foi utilizada a metodologia pesquisa-ação, conduzida junto aos servidores do Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) do Ministério Público de Estado de Sergipe (MPSE), local onde será implantada a ferramenta Galactus Snitch.

Uma vez já disponíveis as principais tipologias de risco de fraude necessárias para a identificação de crimes contra a administração pública, somando-se ao mapeamento sistemático desenvolvido (Souza et al, 2023), a ferramenta foi concebida. A seção 4.5

descreverá detalhes da sua implementação. Ato contínuo, a fim de avaliar a ferramenta proposta, um *survey* foi aplicado junto aos servidores do GAECO do MPSE.

1.6 Organização da Dissertação

Este documento está organizado de acordo com a Instrução Normativa Nº 05/2019/PROCC, a qual permite que a Dissertação seja "uma compilação de artigos científicos submetidos ou publicados em veículos com *Qualis*, desde que seja contextualizada com seções de Introdução, Discussão, Conclusão e Referências, não limitada a estas". São 6 capítulos que fornecem uma base conceitual para o entendimento sistêmico. Os tópicos a seguir descrevem o conteúdo de cada um dos capítulos:

- O Capítulo 1 apresenta esta Introdução, explicando as justificativas, juntamente com os problemas levantados e a suposição de pesquisa;
- O Capítulo 2 apresenta o referencial teórico do trabalho;
- O Capítulo 3 apresenta parte de um Mapeamento Sistemático aceito e publicado na 18th *Conference on Computer Science and Intelligence Systems* - FedCSIS 2023, realizado em Warsaw - Polônia;
- O Capítulo 4 apresenta parte de um artigo submetido à Revista Brasileira de Ciências Criminais, resumindo todo o trabalho efetuado nesta pesquisa. São descritos Planejamento, Operação e Resultados da Pesquisa-Ação;
- O Capítulo 5 traz uma síntese narrativa do Mapeamento Sistemático, juntamente com uma discussão do processo de construção e avaliação da ferramenta;
- Finalmente, no capítulo 6, é apresentada uma compilação de conclusões, contribuições e sugestões de trabalhos futuros.

2.0 REFERENCIAL TEÓRICO

Neste capítulo, é apresentado o referencial teórico necessário para fundamentar a realização deste trabalho. A seguir serão conceituadas as tipologias de risco de fraude e *redflags*, necessárias para o entendimento da forma com que possíveis crimes são sinalizados, seguido do conceito de pesquisa-ação, metodologia sobre a qual o trabalho foi dirigido.

2.1 TIPOLOGIAS DE RISCO DE FRAUDE E *REDFLAGS*

As tipologias de risco de fraude referem-se a categorias ou classificações de atividades, situações ou práticas suscetíveis a atos fraudulentos ou corruptos (Silva, 2022). Elas representam padrões recorrentes e cenários comuns nos quais ocorrem irregularidades financeiras, desvios de recursos ou outras atividades fraudulentas. Essas tipologias são desenvolvidas a partir da análise de casos de corrupção anteriores, experiências passadas e conhecimento especializado em fraude e corrupção. As tipologias de risco de fraude são categorias de atividades suscetíveis a práticas corruptas, que podem variar amplamente, desde desvios de recursos públicos até subornos e nepotismo. Identificar essas tipologias é fundamental para orientar esforços de controle e vigilância.

As *redflags*, por sua vez, são indicadores específicos que levantam suspeitas de possíveis atividades fraudulentas. Eles podem incluir, por exemplo, transações financeiras atípicas, discrepâncias nos registros contábeis e padrões de gastos irregulares. (Silva, 2022) destaca a importância de utilizar a análise de dados e tecnologias avançadas, como a mineração de dados, para identificar essas *redflags* de forma eficaz.

Ao compreender as tipologias de risco de fraude e estar atento às *redflags*, as organizações e os órgãos governamentais podem fortalecer suas estratégias de combate à corrupção, mitigando os danos ao patrimônio público e promovendo maior transparência e responsabilidade na gestão de recursos governamentais.

2.2 PESQUISA-AÇÃO

Essa metodologia tem como objetivo ampliar o conhecimento científico por meio da prática sistemática e da investigação (Filippo, 2010) (Tripp, 2005). O duplo objetivo tem como premissa aprender a partir das ações realizadas dentro de um cenário real, levando o pesquisador a um estágio em que, ao mesmo tempo que tenta compreender os problemas,

deve propor soluções. É importante salientar o aspecto científico das soluções, que devem ser baseadas em teorias e também metrificadas (Filippo, 2010).

Portanto, a pesquisa-ação se distingue tanto de uma pesquisa científica quanto de um estudo prático. Os esforços exigem o desenvolvimento mútuo das áreas de prática e pesquisa. O seu relacionamento recíproco tem como objetivo a melhora do modelo em execução, sendo aperfeiçoado, por meio da pesquisa, continuamente, até alcançar resultados adequados (Tripp, 2005).

O protocolo compreende cinco estágios (Figura 1) que são executados ciclicamente: Diagnosticar, Planejar, Intervir, Avaliar e Refletir. Os estágios serão repetidos progressivamente e indefinidamente até encontrar uma solução satisfatória (Filippo, 2010). Dentro dos estágios, as fases são trabalhadas em um modelo colaborativo, no qual todos os interessados são envolvidos. Pesquisadores e demais profissionais debatem interativamente os problemas, planejam e avaliam em conjunto as soluções (Filippo, 2010). Suas características, análogas ao de ambientes colaborativos, atendem bem ao processo de desenvolvimento de ambientes democráticos digitais como o deste trabalho, pois foca na solução de problemas reais; envolve diferentes grupos interessados; e apoia um desenvolvimento iterativo (Filippo, 2010).

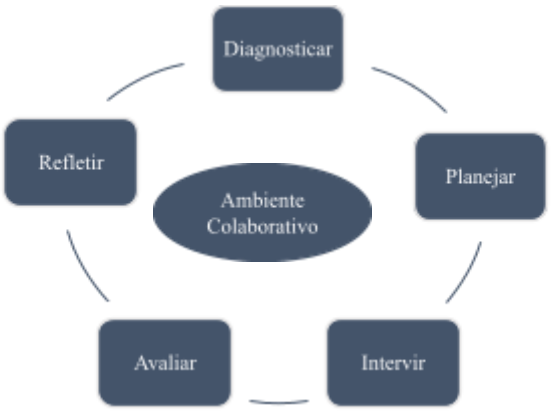

Figura 1: Ciclo de pesquisa-ação para construção do ambiente (Filippo et al, 2010).

3.0 MAPEAMENTO SISTEMÁTICO

Neste capítulo, será apresentada parte do artigo intitulado: Governança Colaborativa em Instituições de Ensino: Uma Revisão Quasi-Sistemática da Literatura. O mapeamento sistemático realizado compôs o primeiro ciclo da pesquisa-ação que será descrita nos capítulos a seguir.

Resumo: Contexto. O gerenciamento de recursos públicos está sujeito a atos ilegais e a identificação automática desses atos depende da análise de um grande volume de dados. **Objetivo.** O objetivo deste trabalho é analisar publicações científicas através de um estudo baseado em um mapeamento sistemático com o propósito de avaliá-las com relação ao uso de ferramentas automatizadas a fim de identificar crimes contra a administração pública em bases de dados a partir da perspectiva de pesquisadores no contexto de ciências de dados. **Metodologia.** Utilizando a estratégia PICO (*Population, Intervention, Comparison and Outcome*), um mapeamento sistemático foi conduzido a fim de encontrar estudos primários na literatura e coletar evidências para o direcionamento da pesquisa. **Resultado.** Dezenove trabalhos foram encontrados de acordo com os critérios estabelecidos. Quase 80% dos estudos encontrados buscavam identificar algum tipo de fraude em processos licitatórios, obtendo acurácias entre 72% e 99%. A pesquisa também revelou diferentes técnicas para abordar o problema. Considerando todos os trabalhos, as bases de dados mais utilizadas foram dados sobre licitações, ações judiciais, editais e estrutura societária de empresas, respectivamente. **Conclusão.** O trabalho mostrou um recente aumento do interesse na análise de dados públicos em busca de irregularidades. Espera-se que esta análise auxiliará órgãos de controle a elucidar diferentes formas de detectar crimes contra a administração pública de forma automatizada.
Palavras-chave: Crime, Corrupção, Administração Pública, Data Science.

3.1 METODOLOGIA

A seção a seguir descreve a metodologia utilizada na execução do trabalho. Para orientar a formulação das questões de pesquisa e a busca bibliográfica foi utilizada a estratégia PICO (Santos et al, 2007). A estratégia PICO orienta a construção da pergunta de pesquisa e da busca bibliográfica e permite que o pesquisador, ao ter uma dúvida ou questionamento, localize, de modo acurado e rápido, a melhor informação científica disponível. Ela apresenta quatro elementos fundamentais da pesquisa: **P**opulação,

Intervenção, **C**ontrole e Resultados (**O**utcome), que os autores utilizaram para descrever todos os componentes relacionados ao problema identificado e estruturar as questões de pesquisa.

3.1.1 Questões de Pesquisa

A fim de responder a questão de pesquisa "RQ1 - Como são analisados grandes volumes de dados em busca de fraudes contra a administração pública?", as seguintes questões de pesquisa de mapeamento (***M**apping **R**esearch **Q**uestion*) foram estabelecidas:

a) MRQ1. Quais os tipos de crime contra a administração pública mais comumente identificados nesses trabalhos?
b) MRQ2. Quais são as abordagens de data science mais amplamente usadas para detectá-los?
c) MRQ3. Quais as métricas de performance das abordagens?
d) MRQ4. Quais as bases de dados mais utilizadas para a aplicação das abordagens?
e) MRQ5. Quais são os principais periódicos e conferências sobre o tema?
f) MRQ6. Em que anos foram publicados mais artigos nesta área?
g) MRQ7. Quais países têm mais publicações nesta área?

3.1.2 Estratégia de Busca

A pesquisa foi esquematizada segundo a estratégia PICO (Santos et al, 2007), e o resultado é ilustrado na Tabela 1. Em seguida foram estabelecidas palavras-chave, em inglês, para cada categoria. O conjunto resultante é descrito na Tabela 2. O conjunto de palavras-chave foi então refinado, sendo retiradas palavras redundantes e identificando radicais das palavras. O resultado do processo é ilustrado na Tabela 3.

Tabela 1: Categorias de estratégia PICO

Categoria	Descrição
População	Publicações que abordem, de forma direta, a identificação de crimes contra a administração pública
Intervenção	Contexto de aplicações que utilizam abordagens automatizadas para identificação de crimes contra a administração pública
Controle	Aplicações que não utilizam abordagens automatizadas para identificação de crimes contra a administração pública
Resultados	Abordagens automatizadas para identificação de crimes contra a administração pública por meio do uso de computação

Tabela 2: Palavras-chave por categoria

Categoria	Descrição
População	crime detection against public administration, corruption detection, collusion detection, fraud detection, corruption in public sector, fraud detection in public procurement, risk pattern in public sector, cartel detection, corruption risk assessment, offenses against public administration, public ghost employee, public ghost payroll, organized crime, prevarication, public treasury, public procurement, public bidding, government purchasing, bid rigging, public fund, money laundering
Intervenção	data mining, data science, text mining, artificial intelligence, a.i, data crossing, crossing technologies, data combination, data manipulation, machine learning, neural network, deep learning, cluster analysis, algorithm
Controle	-
Resultados	decision support system, dss, knowledge discovery, automated system, automated information system, prototype, online analytical processing, olap, intelligent agent, corruption indicator, predictive, model, predictive analytics, model

Tabela 3: Palavras-chave refinadas por categoria

População	Intervenção	Resultados
crime detect*, collusion detect*, corruption detect*, fraud detect*. offences detect*, cartel detect*, prevarication detect*, ghost payroll detect*, ghost employee detect*, bid rigging, money laundering, corruption risk, public administration, public sector, public procurement, public treasury, public bidding, public employ*, government* purchas*, government* treasury, public fund, risk pattern	data mining, data science, text mining, data crossing, artificial intelligence, crossing technologies, data combination, data manipulation, machine learning, neural network, deep learning, cluster analysis, algorithm	decision support system, dss, knowledge discovery, automated system, prototype, automated information system, online analytical processing, olap, intelligent agent, corruption indicator, approach, predictive model, predictive analytics, model

Na Tabela 4 é apresentada a *string* utilizada para as buscas nas bases. As palavras-chave de população foram subdivididas em três blocos, sendo o primeiro relacionado à ação (detecção e suas correspondências), o segundo relacionado ao objeto buscado (crime, corrupção e suas correspondências), e o terceiro bloco relacionado ao local onde encontrar os objetos buscados (setor público, licitações e suas correspondências). Foram utilizadas as buscas em títulos, resumos e palavras-chave nas bases de pesquisa Scopus, IEEE Xplore Digital Library, Web of Science e ACM Digital Library.

Tabela 4: String de busca genérica

(“detect*” OR “search*” OR “find*” OR “look* for” OR “predict*”) **AND** (“crime” OR “corruption” OR “clue” OR “fraud*” OR “collusion” OR “offense” OR “cartel” OR “malfeasance” OR “prevarication” OR “ghost payroll” OR “ghost employee” OR “bid rigging” OR “irregularity” OR ”money laundering”OR “anomaly” OR “suspicious”) **AND** (“public administration” OR “public sector” OR “public procurement” OR “government* procurement” OR “public treasury” OR (“bidding” AND (“public” OR “government*”)) OR “public employ*” OR “government* purchas*” OR “government* treasury” OR “public fund”)	População
AND (“data mining” OR “data science” OR “text mining” OR “artificial intelligence” OR “data crossing” OR “crossing technologies” OR “data combination” OR “data manipulation” OR “machine learning” OR “neural network” OR “deep learning” OR “cluster analysis” OR “algorithm”)	Intervenção
AND (“decision support system” OR “dss” OR “knowledge discovery” OR “automated system” OR “automated information system” OR “prototype” OR “online analytical processing” OR “olap” OR “intelligent agent” OR “predictive model*” OR “predictive analytics” OR “model” OR “corruption indicator” OR “approach*”)	Resultados

A seguir os Critérios de Inclusão: (1) Trabalhos curtos e completos publicados e disponíveis integralmente em bases de dados científicas, com título, resumo e palavras-chave disponíveis em língua inglesa; (2) Trabalhos recentes (publicados a partir de 2010), porém que já possuam aprovação pela comunidade científica. (3) Trabalhos que proponham um método, ferramenta ou aplicação para detecção, seleção ou predição de fraudes ou crimes contra a administração pública em bases de dados por meio do uso da Ciência de Dados. O ano limite de 2010 foi determinado por ser imediatamente anterior à implantação da Lei de Acesso à Informação (Lei nº 12.527/2011) (Brasil, 2011), que regulamentou o direito constitucional de acesso dos cidadãos às informações públicas.

A seguir os Critérios de Exclusão: (1) Trabalhos duplicados; (2) Trabalhos restritos; (3) Trabalhos de revisão; (4) Trabalhos que não busquem a detecção de crimes; (5) Trabalhos que busquem detecção ou predição de outros crimes, fora do contexto deste trabalho.

3.1.3 Estratégia de Extração de Informações

Para avaliar a qualidade do trabalho e responder às questões de pesquisa expostas na seção 3.1.1, foi projetado um formulário a ser respondido para cada artigo lido na íntegra.

Segundo (Kitchenham, 2004), os formulários de extração de dados devem ser projetados para coletar todas as informações necessárias para abordar as questões e os critérios de qualidade do estudo. A Tabela 5 apresenta o formulário de extração utilizado nesta pesquisa. Para os atributos Tipos de Crime, Abordagens, Métricas de Performance e Bases de Dados os resultados são multivalorados, ou seja, há a possibilidade de mais de uma resposta do mesmo atributo para cada artigo.

Tabela 5: Formulário de Extração

Categoria	Descrição
Tipo de Crime	Identificação do tipo de crime contra a administração pública o qual o trabalho se propõe a identificar. Parte dessa tarefa foi realizada já no critério de exclusão 5, que buscou a remoção de trabalhos de identificação de crimes fora do contexto da administração pública
Abordagens	Identificação da abordagem de Data Science utilizada na identificação dos crimes
Métricas de Performance	Identificação do critério de avaliação da abordagem segundo experimento dos autores, caso haja
Base de Dados	Identificação das bases de dados, estruturadas ou não, analisadas pelas abordagens

3.2 Resultados e Discussão

As subseções seguintes descrevem o processo de busca e discutem acerca dos resultados. Na subseção 3.2.1 é descrito o tratamento dos resultados e a aplicação dos critérios de exclusão até a formação da base de análise. A subseção 3.2.2 transcorre brevemente sobre cada trabalho selecionado. A partir da subseção 3.2.3, as questões de pesquisa são respondidas com base nos resultados.

3.2.1 Resultados

Uma vez que os trabalhos foram buscados nas bases especificadas utilizando as palavras-chave, a primeira etapa foi a remoção dos trabalhos duplicados, uma vez que se encontravam em mais de uma base. A Figura 2 apresenta um fluxo descrevendo o processo de extração dos artigos desde esta fase até a análise. A soma das buscas nas bases nos retornou um total de 251 trabalhos, número que foi reduzido a 223 após a remoção dos artigos duplicados.

Em seguida, os outros critérios de exclusão foram aplicados. Dois artigos foram removidos por serem de domínio restrito. Após, foi realizada a leitura do título dos trabalhos para identificação daqueles que se tratavam de artigos de revisão. Junto com o título, foi

observado também o resumo dos trabalhos, que nos permitiu remover os artigos que não tinham como objetivo a detecção de fraudes e crimes. Esses três critérios nos permitiram reduzir o número de trabalhos a um total de 100 artigos.

Após a exclusão segundo esses critérios, nos restavam 100 trabalhos que de fato visavam a busca por técnicas e ferramentas de detecção automática de crimes ou fraudes. Entretanto, muitos desses trabalhos não visavam a identificação de crimes necessariamente na esfera da administração pública. Entre os eventos buscados por esses artigos estavam crimes comuns, invasões hackers, fraudes em planos de saúde e até imigração ilegal. Muitas vezes a leitura do título e resumo eram suficientes para esse discernimento, mas frequentemente a introdução dos artigos precisava ser lida para uma definição mais precisa. Finalmente, após a última etapa na aplicação dos critérios de remoção, chegamos a um total de 19 artigos. Todos foram lidos em sua integridade e um breve comentário sobre os mesmos é descrito nas seções seguintes.

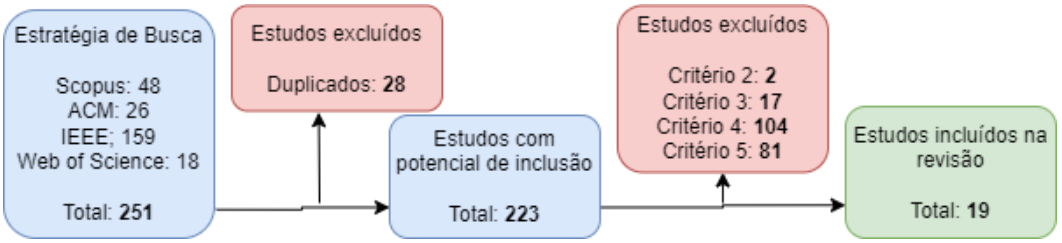

Figura 2: Gráfico de prisma com a extração de dados

3.2.2 Resumos dos Trabalhos

Os trabalhos de (Li et al, 2019) e (Pramanik et al, 2020) apresentam uma abordagem para detecção de crimes baseados em percepções dos usuários. O primeiro baseia-se no conteúdo de postagens na rede social Twitter, enquanto os autores do segundo criaram um survey a ser aplicado por usuários de serviços públicos. O artigo de (Kumar et al, 2020) busca analisar movimentações financeiras a fim de encontrar transações suspeitas que levem ao crime de lavagem de dinheiro, enquanto (Campos et al, 2012) propôs uma ontologia para, aplicado a data warehouse, identificar inconsistências em folhas de pagamento.

A partir daqui, os artigos se concentram na detecção de fraudes na esfera das compras públicas. (Domingos et al, 2016), (Modrusan et al, 2020) e (Rabuzin, 2019) buscam identificar potenciais indícios de fraudes já nos termos de abertura das licitações utilizando, entre outros artifícios, técnicas de text mining. O trabalho de (Carvalho et al, 2014) propôs usar redes bayesianas para identificação de compras fracionadas, onde o processo de licitação

é suprimido caso o valor de cada compra não exceda um valor máximo definido pela legislação brasileira.

O artigo de (Niessen et al, 2020) agregou, junto à base de licitações, uma lista de licitações contra as referidas empresas existem processos na justiça ou reclamações formais. O objetivo é detectar padrões de atributos de processos problemáticos a fim de identificar problemas em novos processos licitatórios usando random forest. (Gallego el al, 2021) utilizaram uma abordagem semelhante, além de utilizar outros dados do processo disponíveis, como orçamento, duração, atrasos, tempo antes de processos eleitorais e padrões geográficos.

Outros trabalhos buscam detectar processos licitatórios com potencial conluio por meio da análise de redes de associação de outras compras, envolvendo os mesmos compradores ou fornecedores. São eles (Ralha e Silva, 2012), (Domashova e Kripak, 2021), (Popa, 2019) e (Carneiro et al, 2020). Para isso, eles utilizam técnicas como regras de associação e random forest. Artigos como (Busu e Busu, 2021), (Padhi e Mohapatra, 2011) e (Davydenko et al, 2017) utilizam algoritmos de clustering para agrupar processos licitatórios entre competitivos e não-competitivos baseados em dados como a razão entre os valores de lance ofertado pelas empresas e o valor inicial do contrato.

Finalmente, (Velasco et al, 2021) e (Martínez-Plumed et al, 2018) propõem a criação de verdadeiras suítes de análise de dados de processos licitatórios, justamente com a adição de informações presentes em outras bases de dados. Isso permite a detecção de esquemas fraudulentos que apenas poderiam ser elucidados a partir da junção dessas informações distribuídas. As bases de dados auxiliares incluem dados de quadro societário das empresas, programas de transferência de renda e dados eleitorais.

3.2.3 **MRQ1 - Quais os tipos de crime contra a administração pública mais comumente identificados nesses trabalhos?**

Em sua grande maioria (78.9%), os autores dos trabalhos centraram seus esforços em técnicas automatizadas para detectar fraudes em licitações, conforme ilustrado na Figura 3. No entanto, os trabalhos diferem sobre o momento em que a tentativa de detecção é realizada. Alguns trabalhos, como (Domingos et al, 2016), (Modrusan et al, 2020) e (Rabuzin et al, 2019), buscam identificar potenciais indícios de fraudes já nos termos de abertura das licitações. Outros trabalhos, como (Busu e Busu, 2021) e (Padhi e Mohapatra, 2011), usam variáveis encontradas durante o processo licitatório para encontrar conluios, como valores de lance e intervalos de tempo. Por fim, trabalhos como (Ralha e Silva, 2012) e (Carneiro et al, 2020) se baseiam na compilação de diferentes processos licitatórios distintos já realizados em

busca de padrões de participação e de vencedores. Há ainda outros trabalhos, como (Velasco et al, 2021) e (Martínez-Plumed et al, 2018), que usam múltiplas abordagens para detecção.

Outros dois trabalhos (Li et al, 2019) e (Pramanik et al, 2020) não definiram um tipo de crime específico, mas se preocuparam em detectar fraudes de uma forma mais abrangente por meio da coleta de opiniões e percepções dos usuários. Há ainda trabalhos direcionados a encontrar fraudes nas folhas de pagamento de servidores (Campos et al, 2012) e lavagem de dinheiro (Kumar et al, 2020).

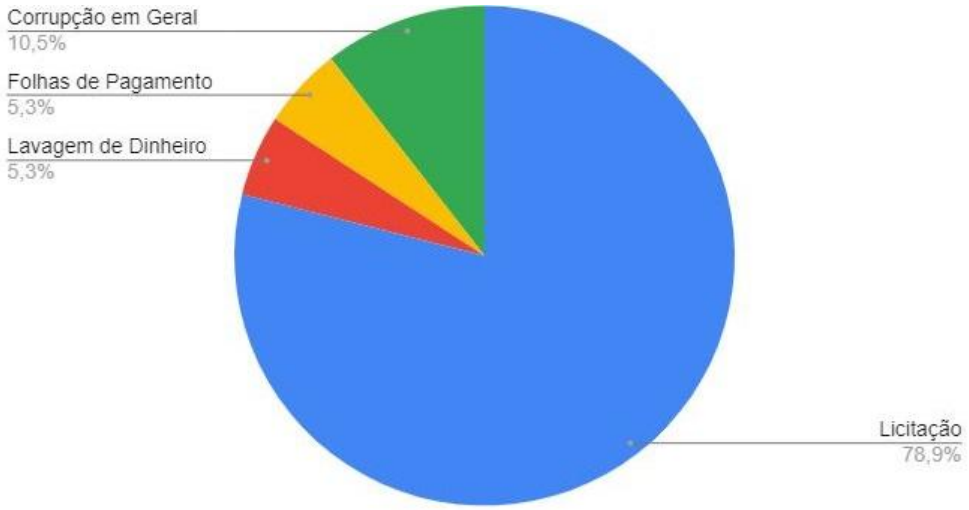

Figura 3: Crimes ou fraudes identificados nas abordagens

3.2.4 **MRQ2 - QUAIS SÃO AS ABORDAGENS DE *DATA SCIENCE* MAIS AMPLAMENTE USADAS PARA DETECTÁ-LOS?**

Como visto no item anterior, a maior parte dos trabalhos focam na detecção de fraudes em compras governamentais por meio de processos licitatórios. Alguns trabalhos, como (Domingos et al, 2016), (Modrusan et al, 2020) e (Rabuzin e Modrusan, 2019), buscam identificar sinais já nos termos de abertura de processo. Para isso, se usam de ferramentas de text mining para análise de elementos específicos dos termos. Uma vez encontrados, aplicam algoritmos de regressão logística ou deep learning para detectar falta de competitividade nos termos da licitação, o que poderia apontar para um eventual conluio prévio entre licitante e empresas interessadas. Por outro lado, os trabalhos voltados à identificação de fraudes nos mesmos processos utilizando dados gerados durante o processo licitatório, como (Busu e Busu, 2021), (Davydenko et al, 2017) e (Padhi e Mohapatra, 2011), usando dados como razão entre valor da oferta e valor inicial da proposta, por meio de algoritmos de clustering para diferenciar processos competitivos e não-competitivos. Por fim,

algoritmos de clustering, bem como regras de associação, são usados por trabalhos como (Ralha e Silva, 2012) e (Carneiro et al, 2020) para identificar conluios entre empresas e fornecedores por meio da análise de vários processos licitatórios em conjunto. Outros trabalhos, como (Velasco et al, 2021) e (Martínez-Plumed et al, 2018), combinam outras técnicas para essa detecção, além da possibilidade de atribuir *scores* para determinadas empresas que participam de processos de licitação. Para isso, fazem uso de outras fontes de dados além da base de contratos e licitações públicas geralmente usadas em outras abordagens, conforme detalhado na seção a seguir.

Os trabalhos (Li et al, 2019) e (Pramanik et al, 2020), que não definiram um tipo de crime específico pois se preocupam em detectar fraudes de uma forma mais abrangente, fazem uso de relatos e impressões dos usuários de serviços públicos. Enquanto (Li et al, 2019) utilizam técnicas de aprendizado de máquina para detectar indícios de fraude nos serviços públicos por meio de postagens no Twitter, (Pramanik et al, 2020) aplicaram formulários para usuários de diversos serviços a fim de buscar indícios de ineficiência baseado nas respostas a esses formulários usando algoritmos de clustering.

Para detectar crimes de lavagem de dinheiro, (Kumar et al, 2020) utilizou um classificador bayesiano baseado em um conjunto de operações bancárias. Já para buscar inconsistências em folhas de pagamento (não necessariamente fraudes) (Campos et al, 2012) definiu um processo de indexação de ontologias por meio de mapas conceituais e indicadores de auditoria como instrumento de documentação de evidências.

3.2.5 **MRQ3 - Quais são as métricas de performance das abordagens?**

De forma geral, os autores utilizaram a acurácia como forma predominante de avaliação estatística dos modelos propostos, à exceção de (Ralha e Silva, 2012) que obteve uma assertividade de 90% segundo um índice próprio de avaliação, denominado RQ, ao tentar identificar formação de cartéis e (Domingos et al, 2016) que obtiveram um Erro Quadrático Médio (EQM) de aproximadamente 0.0013 ao tentar predizer licitações fraudulentas a partir dos seus termos de abertura.

Também ao analisar os termos de abertura das licitações (Modrusan et al, 2020) obtiveram uma a acurácia de 76% utilizando SVM, enquanto (Rabuzin e Modrusan, 2019) obtiveram acurácias entre 72% e 85%, a depender do grupo de produtos do processo licitatório, utilizando Regressão Logística e Redes Bayesianas. (Domashova e Kripak, 2021) obtiveram uma acurácia igual a 67% ao identificar cartéis. Por meio da análise de atributos do

processo licitatório, (Davydenko et al, 2017) chegaram a uma acurácia de 99%, enquanto no trabalho de (Niessen et al, 2020) o mesmo índice foi de 90% utilizando dados semelhantes, incluindo dados de licitações conhecidamente problemáticas. O trabalho de (Carvalho et al, 2014) atingiu uma acurácia de 99.9% analisando licitações fracionadas, onde o valor global é dividido em licitações com valores menores para burlar algumas exigências legais.

Fora do contexto de processos licitatórios, (Pramanik et al, 2020) conseguiu obter uma acurácia de 87.5% na descoberta de irregularidades ao aplicar um questionário a usuários de serviços públicos. (Kumar et al, 2020) atingiu uma acurácia de 81% ao buscar transações financeiras suspeitas em busca de indícios de lavagem de dinheiro. Os outros trabalhos encontrados propuseram modelos de análise de dados sem apresentar validações estatísticas quanto à assertividade desses modelos.

3.2.6 **MRQ4 - Quais são as bases de dados mais utilizadas para a aplicação das abordagens?**

A resposta a essa pergunta deve levar em consideração o tipo de fraude ou crime que o trabalho se propõe a detectar. (Kumar et al, 2020), por exemplo, utilizaram bases de dados de transações bancárias para procurar indícios de fraudes nas mesmas. (Campos et al, 2012) utilizaram uma base de dados de folha de pagamento para montar um data warehouse e definir sua ontologia. Por sua vez, (Li et al, 2019) e (Pramanik et al, 2020) utilizaram postagens na rede social Twitter e dados de um survey aplicado, respectivamente, para identificar fraudes no setor público por meio da percepção dos usuários.

A Figura 4 contabiliza as bases de dados utilizadas para auxiliar a detecção de fraudes em processos licitatórios. Note que uma abordagem pode fazer uso de mais de uma base de dados simultaneamente. Ao todo, 13 dos 15 trabalhos encontrados que propuseram detectar anomalias em processos licitatórios fizeram uso de bases de licitações e compras públicas, enquanto os outros dois analisaram apenas documentos de abertura e definição do processo. A fim de ponderar negativamente processos em que participam empresas contra as quais constavam histórico de processos, alguns trabalhos fizeram uso de bases de processos, sentenças judiciais e reclamações. Outras bases usadas foram as que constavam quadro societário de empresas, programas de transferência de renda, dados eleitorais, dados de pessoas politicamente expostas, dados censitários, registros contábeis e dados cadastrais de empresas.

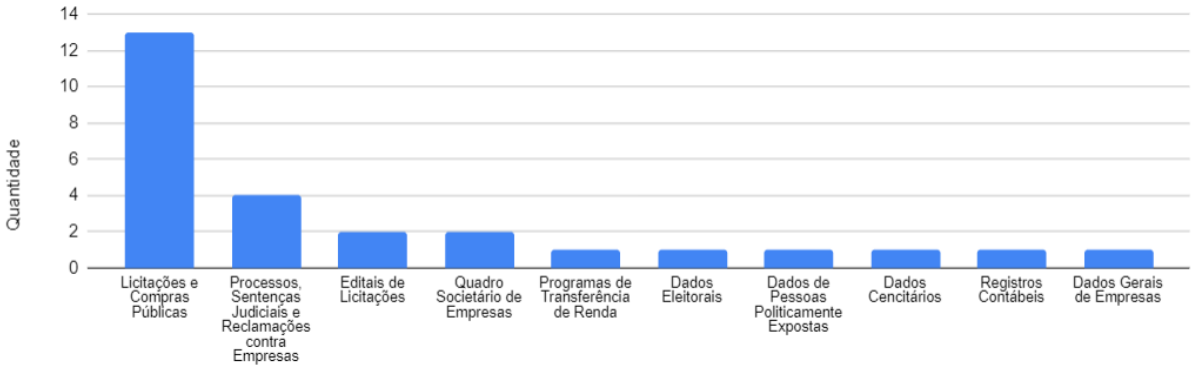

Figura 4: Bases de dados utilizadas para detecção de fraudes em licitação

3.2.7 MRQ5 - Quais são os principais periódicos e conferências sobre o tema?

Dentre os resultados encontrados, todos foram publicados em Revistas, Journals ou Conferências diferentes, não havendo dessa forma um periódico ou conferência que se destacasse perante os demais.

3.2.8 MRQ6 - Em quais anos foram publicados mais artigos nesta área?

Conforme Figura 5, é possível perceber um aumento da publicação de trabalhos que abordam a temática pesquisada a partir do ano de 2019, com 4 trabalhos publicados. O ano de 2020 foi, até então, aquele com o maior número de publicações, contabilizando 6 trabalhos.

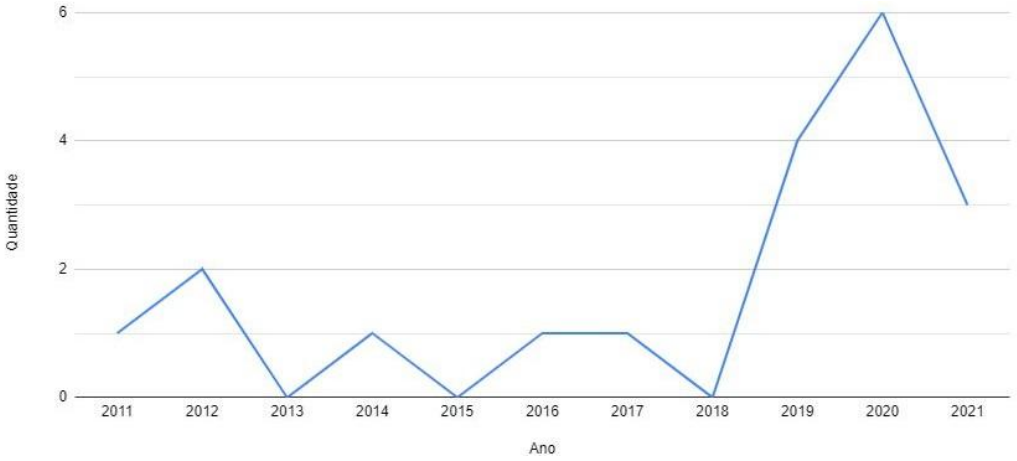

Figura 5: Quantidade de publicações por ano

3.2.9 MRQ7 - Quais países possuem mais publicações nesta área?

A Figura 6 mostra a distribuição de trabalhos publicados ao redor do mundo, onde o tamanho do círculo azul representa a quantidade de publicações. É possível notar que o país

com o maior número de publicações é o Brasil (5), seguido por Rússia (2), Índia (2) e Croácia (2). Colômbia, Espanha, Bangladesh, Reino Unido, Estados Unidos, Romênia, Paraguai e Portugal completam a lista com um trabalho cada. A Figura 5 também plota, em vermelho, a posição do referido país no Índice de Percepção de Corrupção elaborado pela Transparência Internacional referente ao ano de 2020 (Corruption Perceptions Index, 2020), onde quanto mais escura é a cor vermelha, maior a percepção de corrupção. Por meio dos resultados, não foi possível estabelecer relação entre o número de trabalhos publicados e a percepção de corrupção no país.

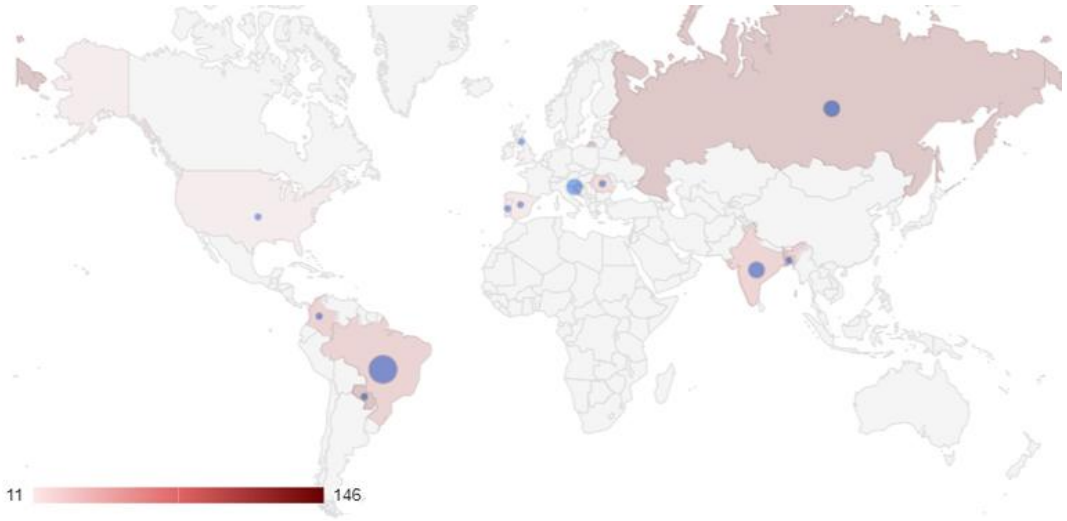

Figura 6: Quantidade de publicações por país e posição no IPC-2020

3.3 Ameaças à Validade

A grande dificuldade do atual trabalho diz respeito à seleção de palavras-chave para busca nas bases. Por mais que o contexto da busca seja bem delimitado, as expressões usadas para descrever crimes, fraudes ou anomalias são diversos, e estão sujeitos a regionalidades e descrições diferentes a depender do país onde as leis são escritas, dificultando a seleção de termos usados como palavras-chave de população, conforme modelo PICO. Tal característica ameaça sobretudo o volume excessivo de trabalhos oriundos do Brasil. Outra dificuldade semelhante é a grande variedade de termos usados para descrever os métodos usados para detecção, descritos nas palavras-chave de intervenção. Uma seleção incompleta de palavras-chave pode limitar consideravelmente o número de resultados retornados.

Quanto aos critérios de exclusão, a heterogeneidade de legislações e políticas distintas dos diversos países pode comprometer a interpretação do pesquisador, tendo em vista

a análise muitas vezes subjetiva desses critérios. Por exemplo, no trabalho atual, as fraudes contra planos de saúde não foram consideradas, dado que o Brasil possui um sistema único de saúde pública pouco entrelaçado com o sistema privado, de modo que as fraudes financeiras contra planos de saúde no Brasil geralmente não envolvem a administração pública. Mas não é possível inferir que isso não ocorra em outros países.

3.4 Conclusão

O trabalho atual descreveu um mapeamento sistemático tendo como objetivo a elucidação de trabalhos científicos que visem o desenvolvimento ou aprimoramento de ferramentas automatizadas para detecção de fraudes ou crimes contra a administração pública. As questões de pesquisa foram levantadas e, baseado na estratégia PICO, as palavras-chave para busca foram selecionadas. Uma vez buscados, os trabalhos foram selecionados baseados em critérios de inclusão e exclusão pré-definidos.

A análise dos resultados mostra que essa preocupação, felizmente, é crescente. Foram vários os trabalhos encontrados visando esse objetivo, e que o fazem abordando diferentes estratégias. Devido ao volume de recursos financeiros envolvidos, as fraudes em licitações são o principal alvo desse tipo de iniciativa. Alguns trabalhos buscam, inclusive, a associação de bases de dados distintas, buscando a descoberta de informações a princípio fragmentadas. Os recursos computacionais para isso vão desde mineração de texto a algoritmos de aprendizado de máquina.

Espera-se que este trabalho possa fornecer orientação para entidades que busquem desenvolver iniciativas e elaborar ferramentas que permitam um melhor acompanhamento das despesas públicas. Conforme observado no trabalho, parte das informações disponíveis para essa tarefa é de domínio público, permitindo que entidades não-governamentais possam participar diretamente dessas iniciativas. Entretanto, são os órgãos de controle que detém domínio exclusivo de parte dos dados identificados como fonte para o trabalho de detecção de crimes, além de possuírem a responsabilidade civil para tal.

É recomendado a existência de trabalhos periódicos nesse sentido, a fim de manter a população e órgãos de controle sempre atualizados sobre as melhores práticas para atingir o objetivo final, que é o combate às fraudes na administração pública. Em trabalhos futuros, recomenda-se um melhor aprofundamento dos termos usados para definir atos ilícitos ou suspeitos e que serão usados na string de busca, a fim de evitar a existência de falsos negativos no processo. Além disso, recomenda-se também uma análise mais aprofundada dos

resultados oferecidos pelas aplicações encontradas, comparando-as e indicando a melhor abordagem para cada situação.

4.0 Criação e Avaliação do Ambiente de Análise de Indícios de fraudes

Neste capítulo, será apresentada parte do artigo intitulado: Criação e avaliação de um ambiente de análise de indícios de fraudes em contratos públicos.

Resumo: Contexto: O gerenciamento de recursos públicos está sujeito a atos ilegais, que privam a sociedade de usufruir desses recursos em sua integralidade. A identificação de tais atos, por sua vez, depende da análise de uma grande quantidade de dados. **Objetivo:** O objetivo do presente trabalho é a criação e avaliação de uma ferramenta capaz de reduzir o escopo de busca por irregularidades, por meio de sinalizações de possíveis fraudes (*redflags*) em contratos firmados pelo estado, no contexto dos servidores do GAECO do MPSE. **Método**: Utilizando a metodologia de pesquisa-ação, uma lista de tipologias de risco de fraude foi identificada, e o Galactus Snitch foi desenvolvido e avaliado por meio de uma pesquisa qualitativa. **Resultados:** o *survey* aplicado revelou uma alta aceitabilidade do Galactus Snitch, onde 66.7% dos servidores afirmaram concordar totalmente que o sistema é capaz de reduzir em mais de 90% o escopo de busca por irregularidades ao apontar possíveis esquemas fraudulentos. 91.7% dos servidores afirmaram também concordar totalmente que as tipologias identificadas são de grande relevância para que o sistema atinja o objetivo esperado. **Conclusões:** utilizando análise de dados de diferentes fontes de informação, o trabalho apresenta uma importante contribuição para o processo de identificação de esquemas fraudulentos em contratos públicos à medida que, segundo os entrevistados, é capaz de reduzir o escopo de busca por fraudes.

Palavras-chave: *Data Science*, Inteligência Artificial, Corrupção, Tipologia Criminal, *Redflag*.

4.1 Metodologia

O estudo foi conduzido com uma abordagem metodológica de natureza aplicada, uma vez que seu foco reside na implementação do conhecimento produzido. A pesquisa aplicada se destaca pela ênfase na aplicabilidade, na utilização e na concretização de ações práticas com base no conhecimento (Gil, 2008).

Seu objetivo apresenta caráter exploratório e descritivo. Uma pesquisa exploratória tem como finalidade a familiarização do problema por meio da análise de dados

ou de observações empíricas (Lakatos e Marconi, 2010). Quanto ao objetivo descritivo, sua conduta procura a caracterização e a determinação de fenômenos ou populações (Gil, 2008).

A pesquisa teve como ponto de partida o Mapeamento Sistemático realizado e publicado em (Souza et al, 2023). A fim de complementar o conhecimento necessário sobre tipologias de risco de fraude comuns a serem identificadas pela ferramenta, e como forma de potencializar os resultados considerando a natureza aplicada, a implementação do trabalho teve também como insumo o *Guia de Combate a Cartéis em Licitação*, publicado pelo Conselho Administrativo de Defesa Econômica (CADE) (CADE, 2019).

Do ponto de vista da pesquisa aplicada, e como forma de avaliar e melhorar a plataforma, foi utilizada a metodologia pesquisa-ação. Esse método foi escolhido por apresentar caráter colaborativo e integrar prática e teoria, sendo, portanto, utilizado como forma de avaliação ativa de uma determinada área ou fenômeno (Filippo, 2010). A pesquisa-ação conduzida permitiu, inclusive, enriquecer a lista de tipologias de risco de fraude identificadas através de entrevistas junto ao GAECO do MPSE. Os ciclos da pesquisa-ação realizada serão detalhados na seção 4.3.

Uma vez já disponíveis as principais tipologias de risco de fraude necessárias para a identificação de crimes contra a administração pública, somando-se ao mapeamento sistemático desenvolvido (Souza et al, 2023), a ferramenta foi desenvolvida. A seção 4.5 descreverá detalhes da sua implementação. Após, um *survey*, cujos resultados serão detalhados na seção 4.6, foi aplicado junto aos servidores do GAECO do MPSE a fim de validar a ferramenta proposta.

4.2 Definição e Planejamento da Pesquisa-ação

Nesta e nas duas próximas seções, este trabalho é apresentado como uma pesquisa-ação.

4.2.1 Definição do Objetivo

Este trabalho tem como objetivo geral a concepção e avaliação qualitativa de um ambiente de análise de indícios de fraudes em contratos públicos, o Galactus Snitch, viabilizando a análise sistematizada de vários contratos firmados entre o poder público e entes privados em busca de possíveis irregularidades.

4.2.2 Planejamento

Nesta seção, é apresentado todo o planejamento inicial realizado para a execução da pesquisa-ação.

4.2.2.1 Seleção de contexto

O contexto escolhido para realização da pesquisa-ação foram os contratos firmados entre entes privados e os órgãos públicos do Estado de Sergipe, aos quais se incluem as prefeituras dos municípios do Estado de Sergipe, bem como o Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) do Ministério Público de Estado de Sergipe (MPSE). O órgão possui como principal objetivo atuar na área de prevenção e repressão às organizações criminosas, crimes contra a ordem econômica e tributária e de investigação criminal no contexto do Estado de Sergipe.

4.2.2.2 Formulação de Questões

Neste ponto é oportuno salientar que a intenção desta pesquisa não é a substituição de agentes humanos de investigação por agentes autônomos. A natureza da investigação criminal exige uma análise subjetiva dos fatos que a cercam, e apesar de não ser possível negar que o atual *status* do uso da Inteligência Artificial converge para uma interferência humana cada vez menor em muitos processos cotidianos, o trabalho atual delimita-se a criar e avaliar uma ferramenta capaz de encontrar indícios de fraudes na administração pública em grandes volumes de dados, para posterior análise dos agentes investigativos.

Dessa forma, e em observância às questões da pesquisa em voga, as seguintes questões serão abordadas no trabalho atual:

a) RQ2 – Quais os principais indicativos da possível existência de fraudes contra a administração pública?;
b) RQ3 – É possível que um software possa reduzir em mais de 90% o escopo de busca de fraudes contra a administração pública ao analisar um grande volume de dados?.

4.2.2.3 Seleção dos Participantes

A partir do terceiro ciclo da pesquisa-ação, e a fim de diagnosticar, elucidar tipologias e avaliar a ferramenta, o projeto admitiu 12 servidores e membros do Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) do Ministério Público de Estado de Sergipe (MPSE). Os servidores pertencem a diferentes setores pertencentes ao

GAECO, com funções distintas, sendo estes a Direção e a Coordenação do GAECO, o Núcleo de Extração de Dados, o Núcleo de Inteligência, o Núcleo de Tecnologia da Informação, o Núcleo de Processamento de Sinais e o Laboratório de Tecnologia contra Lavagem de Dinheiro. A descrição detalhada dos participantes é ilustrada na Tabela 6.

Tabela 6: Detalhamento dos participantes

Servidor	Setor/Função	Experiência
A	Direção ou Coordenação	2 anos
B	Direção ou Coordenação	3 anos
C	Análise Contábil e Documental	2 anos
D	Análise Contábil e Documental	4 anos
E	Análise Contábil e Documental	5 anos
F	Análise Contábil e Documental	6 anos
G	Processamento de Sinais Telemáticos	6 anos
H	Extração de Dados em Dispositivos	3 anos
I	Extração de Dados em Dispositivos	6 anos
J	Inteligência	4 anos
K	Inteligência	6 anos
L	Tecnologia da Informação	6 anos

4.2.2.4 Projeto da Pesquisa-Ação

A pesquisa-ação foi conduzida com base no modelo processual definido por Davison *et al.* (FILIPPO, 2010), descrito na seção 2.2. Entretanto, por se tratar de um modelo flexível, os últimos ciclos tiveram algumas das suas etapas suprimidas. Desta forma, para construção e avaliação da ferramenta, foram realizados 6 ciclos de pesquisa-ação, detalhados a partir da seção 4.3.

4.2.2.5 Instrumentação

Seguem abaixo os recursos utilizados:

a) Galactus Snitch: Plataforma de Análise de Indícios de Fraudes em Contratos Públicos;

- Eclipse IDE: ambiente de desenvolvimento Java;

- Weka: biblioteca de algoritmos de aprendizado de máquina para tarefas de mineração de dados;l
- Git: sistema de controle de versão de código;
- GitLAB: plataforma de versionamento de código, baseada em Git, que permite a criação de repositórios remotos;
- Microsoft SQL Server 2018 Standard Edition: um banco de dados relacional baseado em SQL;
- Notepad++: um editor de texto para código;
- Jitsi Meet: aplicação software livre e de código aberto multiplataforma para voz, videoconferência e mensageiro instantâneo.

b) Questionário de Avaliação de Satisfação para Galactus Snitch:

- Google Forms: aplicativo de gerenciamento de pesquisas. Questionário disponível em: https://forms.gle/v8A36kMyqhygttTX7.

4.2.2.6 *SURVEY*

Conforme mencionado na seção anterior, as avaliações foram realizadas através do uso de um questionário disponível no Google Forms. Antes da avaliação, a solução foi apresentada em reunião conjunta com os mesmos membros e servidores do GAECO do MPSE envolvidos no processo de elicitação de tipologias de risco de fraude. Imediatamente após, o sistema foi liberado para uso para os mesmos membros e servidores. Por fim, após os participantes interagirem com a ferramenta, o *survey* foi aplicado.

Para melhor responder à RQ3, a questão foi desmembrada em Questões de Survey e o questionário possui 4 questões na escala likert-5 e mais 1 pergunta descritiva, totalizando 5 SQs (***S**urvey **Q**uestion*). Em conformidade com o objetivo proposto, as questões abordam a capacidade da ferramenta Galactus Snitch de reduzir o escopo de busca na identificação de contratos fraudulentos, bem como a qualidade das tipologias de risco identificadas durante o processo e sua efetiva aplicação no mundo real. As questões foram listadas a seguir:

SQ1 - Na busca ativa por fraudes em licitações, a aplicação é capaz de reduzir em mais de 90% o escopo de busca por irregularidades, apontando contratos potencialmente fraudulentos?;

SQ2 - Com base nas informações exibidas pela aplicação, já é possível selecionar contratos e licitações para apreciação, a partir dos quais poderiam futuramente resultar na abertura de processo de investigação criminal?;

SQ3 - As *redflags* utilizadas na aplicação possuem alta relevância na tarefa de identificar fraudes em contratos e licitações públicas?;

SQ4 - As *redflags* utilizadas na aplicação são suficientes na tarefa de identificar fraudes em contratos e licitações públicas?;

SQ5 - Em caso de discordância na questão anterior, aponte possíveis *redflags* que poderiam ser usadas para aprimorar a ferramenta (opcional).

Os resultados do *survey* serão descritos na seção 4.6.

4.3 Ciclos de Pesquisa-Ação

Nesta seção, são apresentados os ciclos de pesquisa-ação.

4.3.1 Primeiro Ciclo

O processo de pesquisa-ação foi iniciado com o levantamento bibliográfico do tema, através de um mapeamento sistemático realizado e publicado em (Souza et al, 2023), conforme capítulo 3. O mapeamento objetivou responder primariamente à RQ1 - Como são analisados grandes volumes de dados em busca de fraudes contra a administração pública? Para melhor entender o cenário e especificidades que rodeiam essa questão, a RQ1 foi então dividida em 7 MRQs (*Mapping Research Question*), e o mapeamento sistemático realizado buscou responder às seguintes questões:

a) MRQ1. Quais os tipos de crime contra a administração pública mais comumente identificados nesses trabalhos?;

b) MRQ2. Quais são as abordagens de *data science* mais amplamente usadas para detectá-los?;

c) MRQ3. Quais as métricas de performance das abordagens?;

d) MRQ4. Quais as bases de dados mais utilizadas para a aplicação das abordagens?;

e) MRQ5. Quais são os principais periódicos e conferências sobre o tema?;

f) MRQ6. Em que anos foram publicados mais artigos nesta área?;

g) MRQ7. Quais países têm mais publicações nesta área?

A resposta à MRQ1 identificaria uma tendência que nortearia o restante do trabalho. De acordo com o mapeamento realizado, 78.9% dos trabalhos encontrados

descreviam técnicas de identificação de fraudes em compras públicas, em detrimento a outras fontes de fraudes na administração pública, como folha de pagamento. Alguns trabalhos buscavam identificar potenciais indícios de fraudes já nos termos de abertura das licitações. Outros trabalhos usavam variáveis encontradas durante o processo licitatório para encontrar conluios, como valores de lance e intervalos de tempo. Por fim, houve ainda trabalhos que se basearam na compilação de diferentes processos licitatórios distintos já realizados em busca de padrões de participação e de vencedores.

De uma forma geral, os autores justificaram o interesse nesse tipo de conluio devido aos grandes volumes financeiros envolvidos nas compras governamentais que procedem aos processos licitatórios. Ademais, a quantidade de dinheiro envolvida nessas transações acabam inevitavelmente despertando o interesse de pessoas mal intencionadas.

Outra questão importante a ser respondida eram os tipos de bases de dados utilizadas. Devido ao fato de que a maioria dos trabalhos encontrados buscavam encontrar fraudes em licitações e contratos públicos, as bases de dados que contenham tais informações foi naturalmente a principal fonte de dados utilizada. Assim como esse dado, as informações sobre as outras bases de dados utilizadas, bem como as abordagens de *data science* e métricas de performance encontradas, formaram um bom insumo para os ciclos da pesquisa-ação que viriam a seguir.

4.3.2 Segundo Ciclo

Tendo em vista os resultados do ciclo anterior, e como parte integrante da pesquisa de natureza exploratória, o segundo ciclo da pesquisa-ação buscou fortalecer a coleção de tipologias de risco de fraude já obtida do primeiro ciclo. Para isso, foi consultado o *Guia de Combate a Cartéis em Licitação*, publicado pelo Conselho Administrativo de Defesa Econômica (CADE) (CADE, 2019). O Guia busca consolidar a experiência institucional adquirida pela autoridade brasileira da concorrência nos mais de vinte anos de atuação no combate a cartéis, com destaque para conluios em procedimentos licitatórios. Nesse sentido, o Guia tem como principal objetivo orientar e auxiliar aqueles que participam da elaboração e execução de processos licitatórios, bem como outras autoridades encarregadas da investigação e punição deste ilícito e ao público em geral, sobre como identificar os principais sinais da existência de conduta colusiva em licitações, para assim aprimorar a detecção, prevenção e punição deste ilícito (CADE, 2019).

Conforme esperado, por meio da análise do *Guia de Combate a Cartéis em Licitação*, publicado pelo Conselho Administrativo de Defesa Econômica, a lista de possíveis

tipologias de risco de fraudes foi enriquecida. Algumas dessas tipologias constavam, inclusive, como parte dos elementos observados pelos trabalhos identificados no ciclo anterior. A fim de consolidar e enriquecer as tipologias de risco identificadas, a lista completa foi exposta à análise dos membros e servidores do GAECO do MPSE, no terceiro ciclo da pesquisa-ação.

4.3.3 TERCEIRO CICLO

Os dois primeiros ciclos da pesquisa forneceram, entre outras coisas, um importante extrato de técnicas, métricas e tipologias de fraudes a serem aplicadas no trabalho. No terceiro ciclo, esse conjunto de informações foi então submetido à avaliação de um órgão público responsável pela prevenção e repressão às organizações criminosas no contexto de crimes contra a administração pública. Nessa oportunidade, outras tipologias criminais puderam ser acrescentadas.

O órgão com o qual a pesquisa passou-se a desenvolver a partir do terceiro ciclo foi o Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) do Ministério Público do Estado de Sergipe (MPSE), que possui como principal objetivo atuar na área de prevenção e repressão às organizações criminosas, crimes contra a ordem econômica e tributária e de investigação criminal (Ministério Público do Estado de Sergipe, 2011). O Grupo de Atuação Especial de Combate às Organizações Criminosas possui atuação voltada para obtenção de suporte probatório necessário aos procedimentos, medidas e ações judiciais ou extrajudiciais, através do intercâmbio de informações técnico-jurídicas com órgãos policiais civis, federais ou militares.

O terceiro ciclo se iniciou com uma reunião, na qual estiveram presentes 12 servidores do GAECO do MPSE. Nessa oportunidade, o mapeamento sistemático publicado em (Souza et al, 2023) foi apresentado, assim como os trabalhos encontrados, suas tipologias, bases de dados, técnicas e métricas. Nessa ocasião, também foram apresentadas as tipologias de risco de fraude presentes no *Guia de Combate a Cartéis em Licitação*, publicado pelo Conselho Administrativo de Defesa Econômica (CADE) (CADE, 2019). As tipologias foram validadas, além de outras que foram inseridas pelos próprios servidores nesta etapa. A validação e a inclusão das novas tipologias criminais levaram em consideração sobretudo a experiência profissional dos servidores durante o tempo em que atuaram no órgão.

Até este ponto da pesquisa, somando-se às tipologias identificadas nos dois primeiros ciclos, um conjunto de tipologias criminais foi identificado. Entretanto, nem todas

elas eram passíveis de implementação devido à impossibilidade de acesso a algumas fontes de dados. Essa é uma situação paradoxal, pois a maioria das tipologias dependiam majoritariamente de informações de domínio público, algumas delas se utilizavam de dados cujo acesso dependia de acordos de cooperação técnica entre o MPSE o órgão que os detém.

Desta forma, um subconjunto das tipologias identificadas foi selecionado. Após isso, a primeira versão do sistema foi desenvolvida. A partir de então, o Galactus Snitch foi submetido a outros 3 ciclos de avaliação e construção, que serão descritos na seção a seguir.

4.3.4 Quarto, quinto e sexto ciclos

O quarto, quinto e sexto ciclos ocorreram de forma semelhante. Dado que o terceiro ciclo encerra-se com a implementação da primeira versão do Galactus Snitch, os ciclos a seguir consistem na avaliação da versão atual do sistema, o planejamento da intervenção, e a implementação das mudanças sugeridas. A avaliação foi conduzida por parte dos mesmos servidores presentes no terceiro ciclo, não havendo entre os ciclos processo de avaliação formal. Durante essas intervenções, outras tipologias foram incluídas. A lista de topologias de risco de fraudes implementadas é descrita na seção a seguir. Após o sexto ciclo, o *survey* foi aplicado e os resultados são mostrados na seção 4.6.

4.4 Tipologias de Risco de Fraude

Nesta seção, as tipologias de risco de fraudes identificadas nos ciclos da pesquisa-ação serão descritas. Segundo (Silva, 2022), as tipologias de risco de fraude referem-se a categorias ou classificações de atividades, situações ou práticas suscetíveis a atos fraudulentos ou corruptos. Elas representam padrões recorrentes e cenários comuns nos quais ocorrem irregularidades financeiras, desvios de recursos ou outras atividades fraudulentas. Conforme mencionado na seção 4.3.3, algumas tipologias não puderam ser implementadas devido à impossibilidade de acesso aos dados necessários a sua verificação, e por consequência não serão listadas a seguir.

4.4.1 Procedimento de origem

Os contratos entre o poder público e entes privados podem ser firmados sob alguns tipos de modalidades, como Licitações. Entretanto, há casos onde a licitação é dispensada, conforme artigo 24 da Lei 8.666/93 (Brasil, 1993). Contratos firmados sob

dispensa de licitação possuem um risco de fraude superior aos que seguem todo processo licitatório.

4.4.2 Aditivos e valor aditado

Durante a realização do contrato, por diversas razões, seu valor pode ser majorado bem como seu prazo pode ser estendido. Contratos aditivados possuem um risco associado superior aos demais. É necessário observar a quantidade de aditivos, o valor aditivado bruto, e o valor aditivado em relação ao valor inicial do contrato.

4.4.3 Capital social da empresa

Alguns contratos fraudulentos são firmados com empresas que não possuem capacidade operacional para realizá-los. Uma das formas de verificar essa ameaça é observando a razão entre o valor do contrato e o capital social da empresa que irá executá-lo.

4.4.4 Idade da empresa

Em alguns contratos fraudulentos, as empresas são fictícias e criadas artificialmente, somente para concorrer nas licitações. Empresas executoras muito jovens possuem um risco de fraude superior a empresas com mais tempo de atividade no mercado.

4.4.5 Endereço da empresa

Uma outra forma de verificar se a empresa executora do contrato é fictícia e foi criada artificialmente é verificando seu endereço. Empresas fictícias podem estar registradas em endereços falsos. Muitas vezes esses endereços são residenciais, e pertencentes a pessoas físicas. Um contrato executado por uma empresa cujo endereço é o mesmo do que um endereço registrado para uma pessoa física possui um risco de fraude associado maior.

4.4.6 Sócios beneficiários de programas sociais

Muitas vezes, para evitar suspeitas, as empresas fictícias são registradas em nome de pessoas alheias ao contrato. Muitas dessas pessoas estão condicionadas à situação de vulnerabilidade social, e estão inscritas em programas de transferência de renda do governo. Contratos firmados por empresas cujos sócios são beneficiários do Bolsa Família, por exemplo, possuem um risco associado de fraude maior.

4.4.7 Doação de campanha eleitoral

A doação de campanha, seja ela feita por pessoa física ou jurídica, não é em si ilegal no Brasil. Entretanto, muitas empresas envolvidas em fraudes em contratos públicos

utilizam deste mecanismo como parte do acordo proveniente de conluios nas contratações públicas. Doações feitas pelas empresas executoras de contratos, ou qualquer dos seus sócios, a candidaturas de processos eleitorais de mesmo âmbito do contrato podem sinalizar fraudes, e tais contratos devem ser sinalizados.

4.4.8 Despesa de campanha eleitoral

Da mesma forma da tipologia anterior, a execução de serviços durante a campanha eleitoral, mesmo por empresas que posteriormente venham a ser provedoras de serviço para órgãos públicos, não configura uma irregularidade. Entretanto, tais casos devem ser sinalizados como um possível indicativo de conluio.

4.4.9 Empresa impedida de contratar

Periodicamente o Tribunal de Contas do Estado de Sergipe (TCE-SE) atualiza a lista de empresas impedidas de contratar com o serviço público. Os contratos firmados por essas empresas devem ser sinalizados, ainda que o contrato tenha ocorrido fora do prazo de suspensão determinado pelo próprio TCE-SE.

4.4.10 Empresa citada em denúncia anterior

O GAECO do MPSE já dispõe da lista de denúncias efetuadas em seu portal de denúncias, realizadas pelos cidadãos. Um contrato firmado por uma empresa que foi anteriormente citada em uma dessas denúncias possui um risco superior de existência de fraude.

4.4.11 Padrões de preferência de contratação

A partir da lista de contratos, é possível também tentar estabelecer associações de preferência entre determinados órgãos e empresas, dentro de intervalos de tempo. Para isso, foi utilizado o algoritmo Apriori a fim de minerar regras de associação que buscassem determinar essas preferências.

Apriori é um algoritmo para mineração de conjuntos frequentes de itens e aprendizado de regras de associação em bancos de dados relacionais. Ele identifica os itens individuais frequentes no banco de dados e estende-os a conjuntos de itens cada vez maiores, desde que esses conjuntos de itens apareçam com frequência suficiente no banco de dados. Os conjuntos de itens frequentes determinados pelo algoritmo podem ser usados para determinar regras de associação que destacam tendências gerais na base de dados (Agrawal et al, 1994).

Através do uso do algoritmo Apriori, algumas situações como as citadas abaixo puderam ser identificadas:

a) "Todos os contratos firmados pela empresa XYZ Ltda foram realizados com o órgão Prefeitura Municipal de CDE";
b) "Todos os contratos firmados entre a empresa IJKL S.A. e a Prefeitura Municipal de EFG foram firmados no quadriênio 2017-2020";
c) "95% dos contratos firmados pela empresa OPQ Inc. no quadriênio 2013-2016 foram firmados com a Prefeitura Municipal de UVX".

4.5 Galactus Snitch

O Galactus Snitch é uma ferramenta desenvolvida como fruto desse projeto, numa parceria entre Universidade Federal de Sergipe (UFS) e Ministério Público do Estado de Sergipe (MPSE). Seu principal objetivo é agregar diferentes bancos de dados, a fim de identificar possíveis fraudes contra a administração pública, baseando-se em topologias de fraudes pré-definidas, disponibilizadas visualmente por meio do conceito de *redflags*.

Os principais stakeholders do projeto são promotores de justiça do Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) do MPSE, bem como demais analistas do referido órgão, que possui como principal objetivo atuar na área de prevenção e repressão às organizações criminosas, crimes contra a ordem econômica e tributária e de investigação criminal, no âmbito do Ministério Público estadual (Ministério Público do Estado de Sergipe, 2011). O Grupo de Atuação Especial de Combate às Organizações Criminosas possui atuação voltada para obtenção de suporte probatório necessário aos procedimentos, medidas e ações judiciais ou extrajudiciais, através do intercâmbio de informações técnico-jurídicas com órgãos policiais civis, federais ou militares, cujo acervo será remetido aos órgãos de execução.

O Galactus Snitch surge como um módulo do Galactus, um ambiente projetado, integrador, com capacidade para armazenar todo o acervo digital investigativo de órgãos ou departamentos que atuam com a atividade de Inteligência de Segurança Pública (ISP) (da Cruz Nunes et al, 2019). A infraestrutura do Galactus foi desenvolvida em um ambiente coorporativo de governo, o Ministério Público de Sergipe, com a premissa de integrar, de forma amigável, diversas fontes de dados capazes de trazer aos investigadores, promotores de justiça e demais usuários, agilidade na obtenção das informações. A base do ambiente é um sistema informacional pautado em *web application*, composto por containers de dados

formadores de um DW (Júnior, 2009) que encapsula informações cadastrais e de diversas investigações. Essa integração permite o vínculo automático e interativo de procedimentos investigativos, como, por exemplo, os procedimentos criminais. Os benefícios perpassam a celeridade de atuação e persecução processual penal, provendo à atividade investigativa uma ferramenta de atuação operacional capaz de descobrir conhecimento e gerar novas hipóteses investigativas (da Cruz Nunes et al, 2019).

O Galactus, portanto, agrega uma série de bancos de dados de alta relevância no contexto investigativo criminal, incluindo as fraudes contra a administração pública. As bases de dados presentes nesse ambiente incluem bases de dados públicas, como dados eleitorais oriundos do Tribunal Superior Eleitoral e dados sobre programas de transferência de renda (Bolsa Família). O Galactus incorpora também informações de domínio restrito, adquiridas por meio de convênio entre o Ministério Público do Estado de Sergipe e os órgãos que tutelam essas informações. As bases de dados de domínio restrito incluem informações sobre quadro societário de empresas, endereços de pessoas físicas e jurídicas oriundos de empresas de prestação de serviços básicos, como água e energia elétrica, entre outros.

Dentre as bases de dados que compõem o Galactus, podemos dar destaque às informações de contratos firmados entre o poder público e entes privados. Essas informações foram adquiridas especialmente para atender à demanda do Galactus Snitch, e foram obtidas através de convênio firmado entre o MPSE e Tribunal de Contas do Estado de Sergipe (TCE-SE). É a partir desses contratos firmados que se nortearão as etapas do *pipeline*, onde cada etapa dele implementa a verificação de cada uma das tipologias de risco de fraude identificadas. O esquema geral do processamento das tipologias no Galactus Snitch é mostrado na Figura 7.

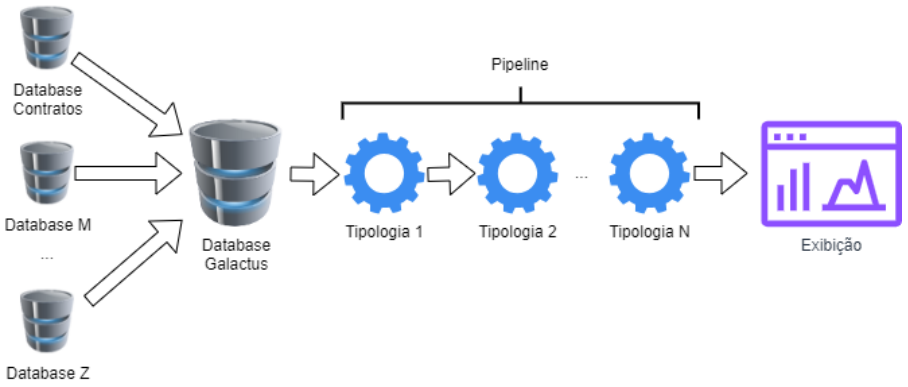

Figura 7: Esquema de processamento das tipologias no Galactus Snitch

Ao final do *pipeline*, a lista de contratos é exibida, bem como a sinalização das *redflags* identificadas com bases nessas tipologias, tendo como insumo os dados provenientes do Galactus. Cada tipologia possui um valor limiar próprio que determina a ativação ou não da sua *redflag*, representada no sistema pela presença da cor vermelha em sua célula correspondente. Há casos também nos quais o valor limiar determina, a partir de valores mínimos e máximos, o nível de transparência da cor vermelha na célula, a fim de representar o grau de criticidade da *redflag*. Os valores limiares para aplicação das *redflags* foram estabelecidos empiricamente, a partir da experiência de investigação dos entrevistados, todavia, podem ser sintonizados de acordo com novas denúncias, mudanças de políticas públicas, leis e mutações contextuais gerais.

Para exibição dos contratos e suas *redflags*, optou-se pelo uso do Qlik Sense Analytics Software, por esta já ser uma ferramenta de uso interno no GAECO do MPSE. O Qlik Sense é uma solução de BI que permite criar análises guiadas orientadas ao usuário, que auxilia na tomada de decisões a partir de fontes diversas de conhecimento, dados, pessoas e ambiente (Qlik, 2023). O acesso ao ambiente é exclusivo a membros e servidores do GAECO do MPSE. A tela do Galactus Snitch onde são exibidos os contratos e suas *redflags* é ilustrada na Figura 8. Para garantir o sigilo de possíveis operações realizadas pelo MPSE, parte das informações exibidas na figura foi suprimida.

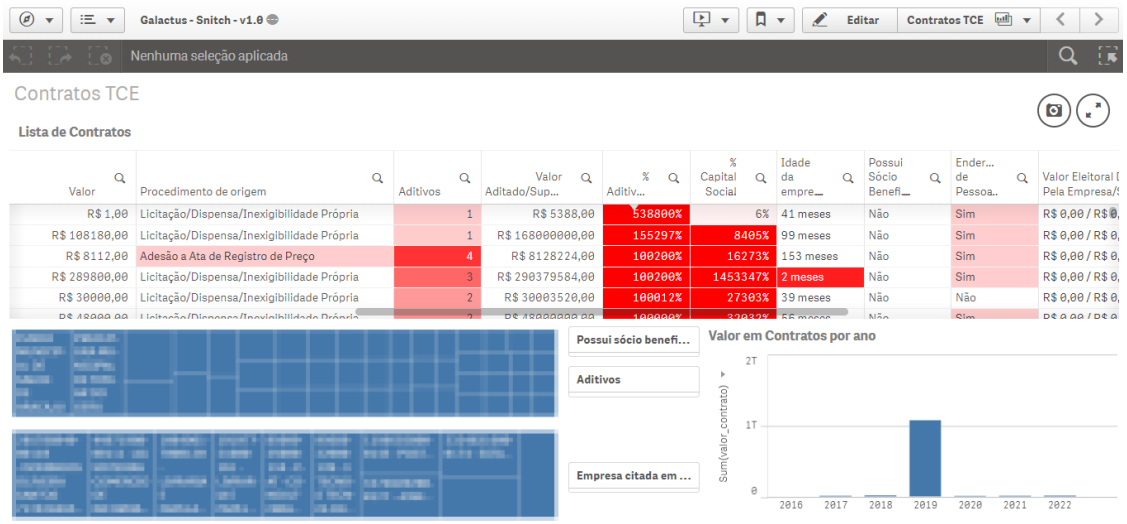

Figura 8: Tela principal do Galactus Snitch

4.6 Resultados

Nesta seção, é apresentada uma análise dos dados coletados e são respondidas as questões de pesquisa.

4.6.1 Análise Sintética dos Dados

Na seção 4.2.2.6, foi estabelecida a RQ3, definida abaixo:

a) O Galactus Snitch é capaz de reduzir em mais de 90% o escopo de busca por fraudes contra a administração pública?

Conforme mencionado na seção 4.2.2.6, a RQ3 foi desmembrada em 5 SQs *(Survey Question*), sendo 4 questões objetivas e 1 questão dissertativa. O questionário foi aplicado a 12 membros e servidores do GAECO do MPSE participantes dos ciclos 3 a 6 da pesquisa-ação, e que tiveram acesso ao sistema Galactus Snitch para avaliação. Os resultados são mostrados a seguir.

4.6.1.1 **SQ1 - Na busca ativa por fraudes em licitações, a aplicação é capaz de reduzir em mais de 90% o escopo de busca por irregularidades, apontando contratos potencialmente fraudulentos?**

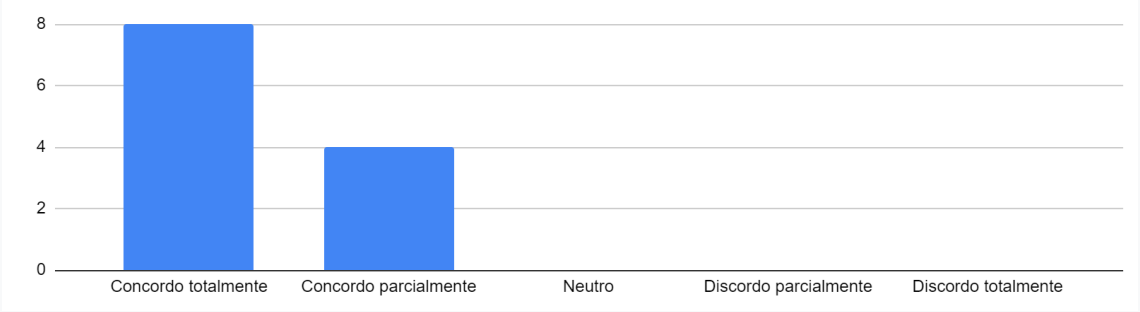

Figura 9: Resposta à SQ1

4.6.1.2 **SQ2 - Com base nas informações exibidas pela aplicação, já é possível selecionar contratos e licitações para apreciação, a partir dos quais poderiam futuramente resultar na abertura de processo de investigação criminal?**

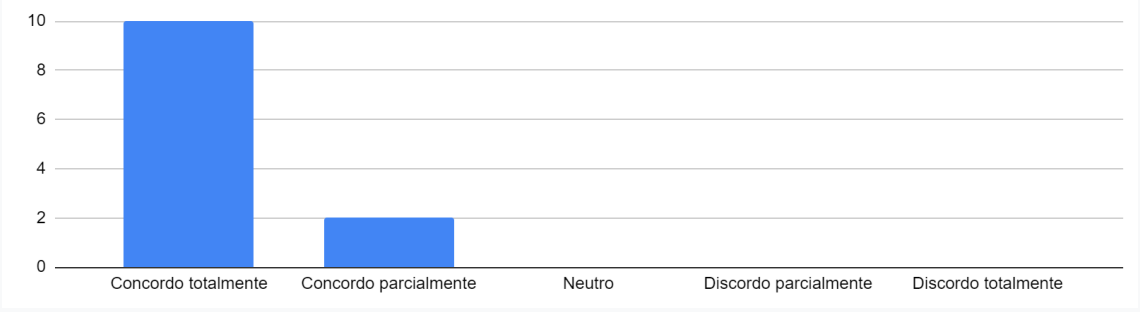

Figura 10: Resposta à SQ2

4.6.1.3 SQ3 - As redflags utilizadas na aplicação possuem alta relevância na tarefa de identificar fraudes em contratos e licitações públicas?

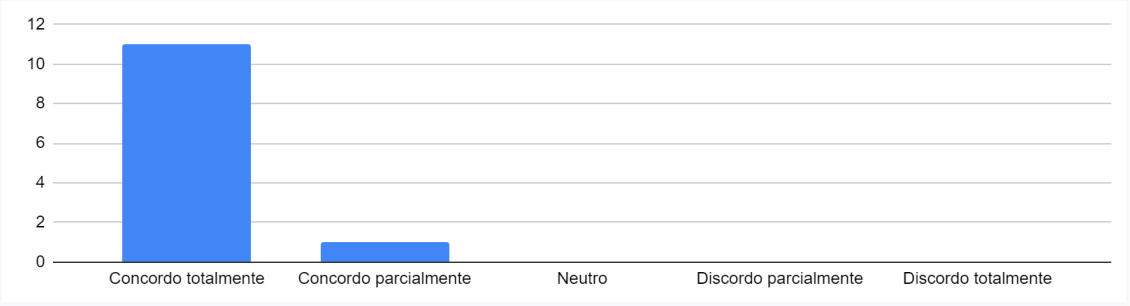

Figura 11: Resposta à SQ3

4.6.1.4 SQ4 - As redflags utilizadas na aplicação são suficientes na tarefa de identificar fraudes em contratos e licitações públicas?

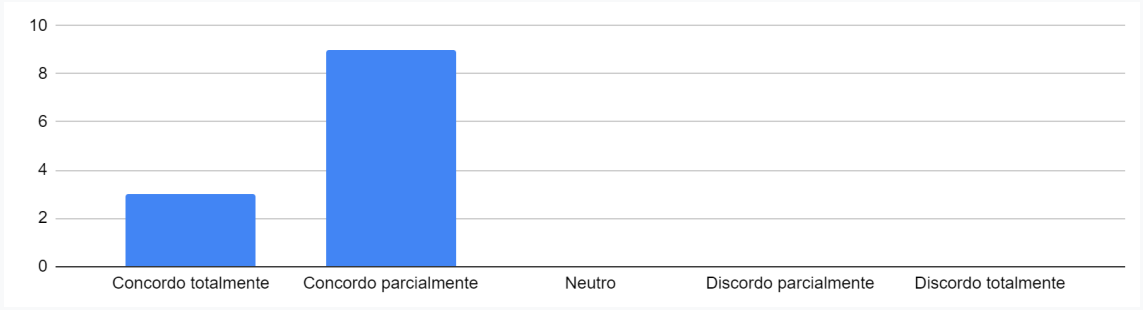

Figura 12: Resposta à SQ4

4.6.1.5 SQ5 - Em caso de discordância na questão anterior, aponte possíveis *redflags* que poderiam ser usadas para aprimorar a ferramenta (opcional).

Para esta questão houve apenas duas respostas, uma vez que a mesma não era de preenchimento obrigatório:

a) "Identificar quem foram as empresas participantes em determinada licitação";

b) "Empresas que sempre ou quase sempre perdem as licitações; Contadores responsáveis por várias empresas; Empresas registradas para o mesmo endereço".

4.6.2 Análise dos Resultados

De uma forma geral, o *survey* aplicado revelou uma alta aceitabilidade do Galactus Snitch no contexto dos servidores do Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) do Ministério Público de Estado de Sergipe (MPSE).

Com relação ao sistema como um todo, 66.7% dos servidores afirmaram concordar totalmente que o sistema é capaz de reduzir em mais de 90% o escopo de busca por irregularidades ao apontar possíveis esquemas fraudulentos, conforme ilustrado na Figura 9, enquanto o restante (33.3%) afirmam concordar parcialmente com a afirmação. Por consequência, 83.3% deles afirmam ainda concordar totalmente que, na versão atual, já é possível selecionar contratos possivelmente fraudulentos para uma análise mais apurada sobre a real existência de irregularidade, enquanto 16.7% afirmam concordar parcialmente com a afirmação, conforme Figura 10.

Ao analisar as tipologias de risco de fraude implementadas, 91.7% dos servidores afirmaram concordar totalmente que elas são de grande relevância para que o sistema atinja o objetivo esperado, enquanto 8.3% dizem concordar parcialmente com essa premissa, conforme Figura 11. Por outro lado, conforme Figura 12, 75% dos servidores afirmaram concordar parcialmente que essas mesmas tipologias são suficientes na tarefa de identificar fraudes em contratos públicos, contra 25% que afirmam concordar totalmente. Esse resultado indica que ainda há margem para expandir a análise de busca de fraudes em contratos públicos, sobretudo agregando ao modelo novas tipologias criminais.

As respostas dissertativas da SQ5 corroboram a conclusão do parágrafo anterior ao mostrar outras possíveis tipologias relevantes nessa análise. Ambas sugestões incluíram, por exemplo, a análise de todas as empresas que participaram do processo licitatório, incluindo as que não venceram. Essa tendência já havia sido observada no mapeamento sistemático realizado (Souza et al, 2023), mas que não foi implementada devido à impossibilidade de acesso aos dados de forma estruturada, conforme mencionado na seção 4.3.3.

4.6.3 Ameaças à Validade

Nesta seção, serão apresentadas as possíveis ameaças à validade identificadas no trabalho.

4.6.3.1 Ameaças à Validade Interna

Com relação ao survey aplicado, é possível citar dois fatores que podem configurar-se como ameaças internas aos resultados obtidos. O primeiro mantém relação à SQ2, na qual a maioria dos servidores afirmaram já ser possível iniciar um procedimento de investigação a partir da versão atual do Galactus Snitch. A ameaça surge da possível falta de materialidade quando de fato o procedimento de investigação for aberto, dado que muitas vezes o sucesso deste procedimento investigatório depende da possibilidade de acesso a dados

cuja a quebra de sigilo bancário, fiscal ou telefônico é mandatória. O acesso a esses dados restritos só é possível após autorização judicial, e uma certa materialidade de um ato fraudulento deve existir para que essa liberação ocorra pela justiça. A mitigação dessa ameaça surgirá da observação prática do uso da ferramenta.

A segunda ameaça interna mantém relação às respostas da SQ3 e SQ4, nas quais a maioria dos servidores afirmam concordar, ainda que parcialmente, que as tipologias implementadas são relevantes e até suficientes na tarefa de se identificar atos fraudulentos em contratos públicos. Em que pese a razão do parágrafo anterior, que a possível suficiência só será comprovada na prática, essa sensação de suficiência pode ter sido causada devido ao fato de que os mesmos servidores que participaram do processo de seleção de parte das tipologias foram os que avaliaram a ferramenta, o que pode ter representado um viés de resposta. A mitigação dessa ameaça passa pela implementação de uma validação cruzada, seja ela interna ao órgão ou que envolva outros órgãos de investigação.

4.6.3.2 Ameaças à Validade Externa

A pesquisa foi conduzida com uma amostragem formada pelo público para o qual o sistema está direcionado: membros e servidores de uma organização que objetiva o combate a organizações criminosas. Entretanto, pelo fato de apenas uma organização ter sido selecionada na condução da pesquisa e avaliação dos resultados, isso pode se tornar uma ameaça ao entendimento de que esse mesmo resultado pode ser estendido para um contexto mais amplo, que envolva outros órgãos de investigação, inclusive de outros estados da federação. Uma avaliação mais ampla poderá ser aplicada caso essa tese deseje ser comprovada.

4.6.3.3 Ameaças à validade de Construção

Para mitigar as ameaças de construção, o projeto foi desenvolvido sob o modelo de pesquisa-ação, seguindo o ciclo de fases estabelecido pela literatura (Filippo, 2011). Para contornar o viés advindo de fatores humanos, não foram apresentados os objetivos e as questões de pesquisa aos participantes.

4.7 Conclusão e Trabalhos Futuros

O presente trabalho apresenta uma importante contribuição para o processo de identificação de esquemas fraudulentos em contratos públicos. Adotando uma abordagem de pesquisa-ação, a pesquisa envolveu as pessoas de interesse do projeto, responsáveis por investigar tais fraudes no contexto da administração pública no Estado de Sergipe.

Evolutivamente, a cada ciclo, as tipologias de risco de fraude relevantes na execução dessa tarefa foram sendo identificadas, um sistema de análise dessas tipologias e exibição das respectivas redflags foi desenvolvido - o Galactus Snitch - e ao final do processo esse mesmo sistema foi avaliado.

As respostas ao questionário, respondidas pelos mesmos 12 membros e servidores do Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) do Ministério Público de Estado de Sergipe que participaram do processo de identificação das tipologias, indicou uma alta e imediata aplicabilidade da ferramenta em contexto prático, destacando-se a alta relevância das tipologias implementadas no Galactus Snitch. Contudo, embora as respostas dos formulários tenham indicado essa alta aplicabilidade, é necessário observar sua condução no contexto prático.

Por fim, como trabalhos futuros, é necessário a expansão das tipologias implementadas, sobretudo envolvendo todas as empresas participantes dos processos licitatórios. Outras tipologias podem surgir a partir do acesso a outras bases, tais como de advogados e contadores das empresas. É necessário também avaliar a aplicabilidade do modelo proposto em outros contextos, como outros órgãos de investigação de crimes da mesma natureza de outros Estados da federação.

5.0 DISCUSSÃO

Neste capítulo, será apresentada uma discussão dos resultados obtidos após a realização do Mapeamento Sistemático e a Construção e Avaliação do Galactus Snitch.

Quantitativamente, a observação dos resultados do Mapeamento Sistemático permitiu observar que a busca por formas automatizadas para detecção de fraudes e crimes é relativamente recente no contexto científico. Para muitos autores, isso se dá muitas vezes pelo processo tardio de digitalização dos governos em relação à iniciativa privada, sobretudo em países subdesenvolvidos. Sem dados governamentais disponíveis no formato digital, não há meios para execução de tal tarefa.

A análise quantitativa também colocou o Brasil como o grande contribuidor nesse tipo de abordagem, apesar do mesmo estar numa posição intermediária no Índice de Percepção da Corrupção (IPC), no ano de 2020, provido pela Transparência Internacional. Foi observado que o quantitativo de publicações por país não pôde ser relacionado diretamente às percepções de corrupção dos países, segundo o mesmo índice.

Os processos de licitação, conforme análise, são o grande alvo de processos automatizados de detecção de fraude no contexto público. De uma forma geral, os autores justificaram o interesse nesse tipo de conluio devido aos grandes volumes financeiros envolvidos nas compras governamentais que procedem aos processos licitatórios. Ademais, a quantidade de dinheiro envolvida nessas transações acabam inevitavelmente despertando o interesse de pessoas mal intencionadas.

Para tais detecções, os trabalhos se revezam no uso de modelos preditivos e dedutivos. Os modelos dedutivos são geralmente embasados em legislações locais e conhecimento prévio acerca das tipologias de esquemas fraudulentos, o que se revela muitas vezes em uma desvantagem pois esse tipo de abordagem não é capaz de prever novos formatos de esquemas. Por outro lado, modelos preditivos são mais difíceis de aplicar pela ausência de bases de treinamento, tendo em vista que o número de fraudes comprovadas é frequentemente insuficiente para modelagem desse tipo de abordagem.

A análise dos trabalhos encontrados através do mapeamento sistemático permitiu identificar uma grande lacuna no estado da arte: a grande maioria dos trabalhos que buscavam detectar crimes contra a administração pública, mais precisamente no contexto de compras públicas, o faziam a partir da análise de uma única tipologia criminal. Em que pese a importância de trabalhos dessa natureza, faz-se necessária a implementação de um ambiente de análise de indícios de fraudes baseado na implementação de múltiplas tipologias em

conjunto, incluindo, mas não limitando-se a, essas abordagens individuais implementadas nesses trabalhos.

Dessa forma, para a execução do trabalho atual, os esforços iniciais se concentraram na identificação de tipologias de riscos de fraudes no contexto da administração pública. A implementação dessas tipologias está sujeita ao acesso aos dados necessários à sua implementação. Essa limitação de acesso a dados, sobre os quais são regidas legislações específicas de proteção, determinou os Contratos Públicos como o ativo central de apreciação no presente trabalho, dado que esse tipo de informação era disponível ao GAECO do MPSE, local de aplicação da pesquisa. Pode-se entender a análise de contratos públicos como uma extensão à análise de processos licitatórios, tendo em vista que muitos contratos são firmados por empresas selecionadas em processos de licitação. A ausência do processo licitatório no processo de contratação, inclusive, revelaria-se ser uma das tipologias de risco de fraude levantadas no processo de elicitação de tipologias criminais realizado no trabalho. Por outro lado, a ausência de dados específicos de licitações, como os participantes perdedores, limitou o número de tipologias aplicáveis identificadas no mesmo processo.

Seguindo a pesquisa-ação, essas tipologias criminais foram identificadas tendo como base três fontes:

a) O mapeamento sistemático realizado que, como mencionado, identificou trabalhos em sua maioria voltado para a identificação de fraudes em processos licitatórios;
b) O Guia de Combate a Cartéis em Licitação, publicado pelo Conselho Administrativo de Defesa Econômica (CADE);
c) As entrevistas que foram realizadas com os servidores do GAECO do MPSE.

Conforme mencionado, o número de bases de dados disponíveis era limitado, o que acabou cerceando, por consequência, o número de tipologias de risco de fraude passíveis de implementação, dada suas dependências de dados específicos. Uma vez identificadas as tipologias, as mesmas foram filtradas de acordo com a pré-existência das informações necessárias a sua implementação, resultando num subconjunto de tipologias. A base de dados cuja ausência limitou o maior número de tipologias criminais foram as informações sobre empresas participantes de processos licitatórios.

Dentre as principais bases de dados disponíveis, podemos citar: Contratos Públicos, Quadro Societário de Empresas, Dados Cadastrais de Empresas, Dados de Programa de Transferência de Renda (Bolsa Família), Processos Eleitorais, entre outros. Os bancos de dados dividem-se entre bancos de dados públicos e aqueles de domínio restrito obtido através

de convênio. Muitas dessas bases auxiliares eram pré-existentes no âmbito do Grupo de Atuação Especial de Combate ao Crime Organizado (GAECO) do Ministério Público de Sergipe (MPSE) e parte integrante do Galactus - um ambiente projetado, integrador, com capacidade para armazenar todo o acervo digital investigativo de órgãos ou departamentos que atuam com a atividade de Inteligência de Segurança Pública (ISP), e desenvolvido pelo próprio GAECO do MPSE.

Uma vez selecionado o subconjunto de tipologias criminais, elas foram implementadas na forma de um *pipeline* que teve como base os Contratos Públicos. Para cada contrato, uma série de verificações era realizada e seu resultado era anexado como metainformação do próprio contrato. Essas informações são as que determinam os riscos associados àquele contrato, e foram materializadas através das *redflags*. Essas verificações, conforme tipologias implementadas, muitas vezes eram fruto do cruzamento de informações das bases de dados auxiliares pré-existentes.

Após a execução do *pipeline*, os dados dos contratos, juntamente com os metadados adicionados posteriormente, ficam dessa forma disponíveis para consulta através de uma interface web. Para exibição das informações foi escolhida a interface Qlik Sense, pela sua relevância no mercado e licença disponível para uso no âmbito do GAECO do MPSE. O Galactus Snitch estava então disponível para avaliação por parte dos servidores do mesmo órgão.

O *survey* aplicado revelou uma boa aceitabilidade do Galactus Snitch por parte dos servidores. A resposta à SQ1 apontou que todos os servidores concordaram que o sistema é capaz de reduzir em mais de 90% o escopo de busca por irregularidades ao apontar possíveis esquemas fraudulentos, dos quais 66.7% afirmaram concordar totalmente com essa afirmação, enquanto os outros 33,3% afirmaram concordar parcialmente, alegando a ausência de algumas tipologias ainda não implementadas. É importante ressaltar que a avaliação aplicada deu-se de forma qualitativa, de acordo com a percepção de cada avaliador em relação à redução do escopo de busca a uma razão mínima de 90%. À medida que novas investigações forem sendo realizadas, dados precisos poderão ser coletados, a fim de averiguar o percentual exato alcançado.

Quanto à SQ2, a maioria deles (83.3%) afirmaram também já ser possível selecionar contratos para apreciação inicial, e que a partir de então poderia-se iniciar um processo de investigação criminal. Os resultados da SQ1 e SQ2 indicam, portanto, a possibilidade do uso de uma ferramenta capaz de auxiliar na identificação de crimes dessa natureza no contexto da administração pública.

Por outro lado, as respostas às SQ3 e SQ4 revelam a qualidade das tipologias criminais identificadas durante todo o processo. Na resposta à SQ3, 91.7% dos servidores afirmaram concordar totalmente que as tipologias implementadas são de grande relevância para que o sistema atinja o objetivo esperado, enquanto que, em resposta à SQ4, 75% dos servidores afirmaram concordar parcialmente que essas mesmas tipologias são suficientes na tarefa de identificar fraudes em contratos públicos. Tendo em vista que a principal base de dados utilizadas no trabalho atual foi a de Contratos Públicos, que a implementação das tipologias depende primordialmente da disponibilidade das bases, e que o mapeamento sistemático revelou que a principal base usada em trabalhos realizados foi a de Licitação e Compras Públicas, os resultados da SQ3 e SQ4 indicam a capacidade das informações sobre Contratos Públicos substituir a base de Licitações e Compras na execução da tarefa.

Entretanto, em consonância com o mapeamento sistemático, as respostas à SQ5 ainda revelam a necessidade de expandir as tipologias implementadas, umas vez que essas respostas foram unânimes ao sugerir a inclusão na análise de dados de participantes das licitações, em adição à análise apenas dos seus vencedores. O mapeamento realizado revelou trabalhos como (Ralha e Silva, 2012), (Domashova e Kripak, 2021), (Popa, 2019) e (Carneiro et al, 2020), que buscaram detectar processos licitatórios com potencial conluio por meio da análise de redes de associação de outras compras, envolvendo os mesmos compradores ou fornecedores. A inclusão de dados de participantes das licitações pode também fortalecer as tipologias já existentes, como a identificação de empresas fictícias perdedoras (e não somente vencedoras) no processo através da análise da idade da empresa, endereço coincidente com endereços residenciais e sócios inscritos em programas de transferência de renda.

Apresentada a seção de Discussões, no próximo capítulo, serão abordadas as contribuições, conclusões e dificuldades obtidas, bem como os possíveis trabalhos futuros.

6.0 CONCLUSÃO

Em todo o mundo, em maior ou menor grau, a má gestão do dinheiro público retira da população o direito de usufruir integralmente dos recursos que por ela foram providos por meio dos impostos. Essa má gestão muitas vezes é intencional, decorrentes de ações criminosas que visam subtrair ou usar bens públicos em benefício próprio indevidamente. Felizmente, a recente digitalização dos serviços governamentais, aliado ao princípio que parte dessa carga de informação é de domínio coletivo, permite o surgimento de iniciativas, organizacionais ou populares, que visem a identificação desses atos ilícitos. O grande volume dessas informações, entretanto, exige um processo automatizado.

Dado este cenário, este trabalho se propôs a conceber e avaliar qualitativamente um ambiente que integre diferentes bases de dados e sinalize possíveis fraudes contra a administração pública a fim de dar suporte aos órgãos de investigação na busca ativa por crimes nesse contexto. Para possibilitar a realização do objetivo geral, podemos enumerar os seguintes objetivos específicos:

a) OE1 - Execução de um Mapeamento Sistemático do estado da arte de ferramentas e técnicas que busquem a identificação ou sinalização da existência de fraudes no contexto da administração pública;
b) OE2 - Identificação de potenciais tipologias criminais que sinalizem a possível existência de esquemas fraudulentos em contratos públicos;
c) OE3 - Construção e avaliação de um sistema de análise, fruto de pesquisa-ação avaliativa, que integre diferentes bases de dados e sinalize possíveis contratos com fraude através do uso de *redflags*.

Conforme visto no capítulo 3, o objeto OE1 foi atingido com a execução do Mapeamento Sistemático, que detalhou a busca por ferramentas e técnicas que visam a identificação de fraudes nesse contexto a partir do estabelecimento de 7 questões de pesquisa. As respostas a essas questões detalhou o tipo de crime mais comumente identificado, suas principais técnicas, e auxiliou na identificação das tipologias criminais passíveis de implementação a fim de atingir o objetivo OE2.

O *survey* aplicado, conforme capítulo 4, permitiu verificar o cumprimento dos objetivos OE2 e OE3. As questões SQ3 e SQ4 do mesmo survey permitiram avaliar a qualidade e relevância das tipologias identificadas durante o processo (OE2), enquanto as questões SQ1 e SQ2 avaliaram que é possível utilizar o Galactus Snitch como ferramenta de integração de dados e utilização prática na tarefa de identificar proativamente crimes contra a

administração pública, no contexto específico dos contratos públicos firmados (OE3). Finalmente, é importante salientar a dificuldade na obtenção de algumas bases de dados como o principal desafio enfrentado neste trabalho.

6.1 Contribuições

Dando prosseguimento ao tópico anterior, como principais contribuições deste trabalho, destacam-se:

- Mapeamento Sistemático do uso de *Data Science* na identificação de crimes contra a administração pública. Seus resultados foram submetidos e publicados na 18th Conference on Computer Science and Intelligence Systems - FedCSIS 2023, realizado em Warsaw - Polônia;
- O catálogo de diversas tipologias criminais identificadas e validadas, tendo como origem, além do Mapeamento Sistemático, o Guia de Combate a Cartéis em Licitações do Conselho Administrativo de Defesa Econômica (CADE), e entrevistas realizadas em um órgão de investigação criminal - o Gaeco do MPSE;
- O desenvolvimento do Galactus Snitch - uma ferramenta capaz de agregar diferentes bancos de dados, a fim de identificar possíveis fraudes contra a administração pública.

Na próxima seção, serão apresentados possíveis desdobramentos relacionados a este trabalho.

6.2 Trabalhos Futuros

Apesar do cumprimento dos objetivos estabelecidos, podemos elencar algumas lacunas deixadas no presente visando aprimorar a realização do objetivo principal.

Quanto às **tipologias criminais** identificadas, conforme citado no capítulo 5, nem todas puderam ser implementadas e a principal razão deve-se à impossibilidade de acesso a algumas fontes de dados. Sugere-se que, uma vez disponíveis essas informações, a ferramenta seja incrementada com a adição dessas tipologias. A consulta ao mapeamento sistemático realizado poderá nortear sobre as próximas técnicas a serem abordadas. Dentre as principais, podemos citar:

- A identificação de rodízio entre os participantes através da análise de vencedores e perdedores dos processos licitatórios;

- A identificação de empresas, fictícias ou reais, que aderem ao processo licitatório destinadas a serem perdedoras;
- A comparação de agentes procuradores comuns entre empresas participantes de processos de licitação;
- A análise dos padrões de preços definidos pelas empresas nos processos de pregão;
- A identificação de compras fracionadas, resultando na supressão do processo de licitação;
- A análise de condições excessivamente restritivas nos termos de abertura de licitações;

Quanto à **validação** da ferramenta, tendo em vista que no trabalho atual o mesmo órgão de combate à corrupção foi usado como participante do processo de construção e avaliador da ferramenta resultante, sugere-se a inserção de outros órgãos de mesmo objetivo no processo, com intuito de validar as tipologias e a ferramenta de forma cruzada. Essa medida poderia ampliar as tipologias sugeridas, bem como aumentar a confiabilidade do resultado da validação.

Quanto à **confecção** da ferramenta Galactus Snitch, uma vez que a interface de visualização de *redflags* foi desenvolvida como um *dashboard* do Qlik Sense, é preferível que esta disponha de uma interface de visualização própria. Esta medida poderá acarretar em filtros mais personalizados de busca e um melhor controle de acesso por parte dos usuários.

O processo também poderá ser incrementado com o uso de um ***score***. A versão atual da ferramenta exibe as *redflags* em tela, baseado nos riscos tipológicos implementados, tornando a identificação de possíveis fraudes uma tarefa visual. Adicionalmente, essas *redflags* poderiam ser reduzidas a um *score*, através da confecção de uma função matemática que infira pesos coerentes às tipologias e ordene os contratos de acordo com o risco total associado a eles.

REFERÊNCIAS

Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).

Brasil. (1993). "Lei n.º 8.666/1993". Regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da Administração Pública e dá outras providências. *Diário Oficial da República Federativa do Brasil*. https://www.planalto.gov.br/ccivil_03/leis/l8666cons.htm.

Brasil. (2011). "Lei n.º 12.527/2011". *Diário Oficial da República Federativa do Brasil*. https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.

Busu, M., & Busu, C. (2021). Detecting bid-rigging in public procurement. A cluster analysis approach. *Administrative Sciences*, *11*(1), 13.

Campos, S. R., Fernandes, A. A., De Sousa, R. T., De Freitas, E. P., da Costa, J. P. C. L., Serrano, A. M. R., ... & Electrical, C. T. R. (2012, November). Ontologic audit trails mapping for detection of irregularities in payrolls. In *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)* (pp. 339-344). IEEE.

Carneiro, D., Veloso, P., Ventura, A., Palumbo, G., & Costa, J. (2020). Network analysis for fraud detection in portuguese public procurement. In *Intelligent Data Engineering and Automated Learning–IDEAL 2020: 21st International Conference, Guimaraes, Portugal, November 4–6, 2020, Proceedings, Part II 21* (pp. 390-401). Springer International Publishing.

Carvalho, R. N., Sales, L., Da Rocha, H. A., & Mendes, G. L. (2014, July). Using Bayesian Networks to Identify and Prevent Split Purchases in Brazil. In *BMA@ UAI* (pp. 70-78).

Corruption Perceptions Index (2020). Disponível em: https://www.transparency.org/en/cpi/2020. Acessado em: 10/10/2021.

da Cruz Nunes, F. M., Júnior, M. C., da Silva Junior, J. B., Costa, L. B. B., & Recchi, E. C. S. (2019, May). Galactus-Um ambiente inteligente para apoio à tomada de decisão no âmbito do Ministério Público de Sergipe. In *Anais Estendidos do XV Simpósio Brasileiro de Sistemas de Informação* (pp. 153-156). SBC.

Davydenko, V. I., Morozov, N. V., & Burmistrov, M. I. (2017, August). Adaptation of cluster analysis methods in respect to vector space of social network analysis indicators for

revealing suspicious government contracts. In *2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)* (pp. 57-62). IEEE.

Domashova, J., & Kripak, E. (2021). Application of machine learning methods for risk analysis of unfavorable outcome of government procurement procedure in building and grounds maintenance domain. *Procedia Computer Science*, *190*, 171-177.

Domingos, S. L., Carvalho, R. N., Carvalho, R. S., & Ramos, G. N. (2016, December). Identifying it purchases anomalies in the Brazilian Government Procurement System using deep learning. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 722-727). IEEE.

FIESP. (2010). Relatório Corrupção: custos econômicos e propostas de combate. In DECOMTEC. FIESP - Federação das Indústrias do Estado de São Paulo.

Filippo, D. (2011). Pesquisa-ação em sistemas colaborativos. *Sistemas colaborativos*, *1*.

Gallego, J., Rivero, G., & Martínez, J. (2021). Preventing rather than punishing: An early warning model of malfeasance in public procurement. *International Journal of Forecasting*, *37*(1), 360-377.

Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. 6. ed. Editora Atlas SA.

Investopedia. (2021). Credit Score. https://www.investopedia.com/terms/c/credit_score.asp. Acessado em: 16/01/2021.

Júnior, M. C., Mendonça, M., & Rodrigues, F. (2009, October). Data warehousing in an industrial software development environment. In *2009 33rd Annual IEEE Software Engineering Workshop* (pp. 131-135). IEEE.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, *33*(2004), 1-26.

Kumar, A., Das, S., & Tyagi, V. (2020, October). Anti money laundering detection using Naïve Bayes classifier. In *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)* (pp. 568-572). IEEE.

Lakatos, E. M., & Marconi, M. D. A. (2010). Fundamentos da metodologia científica. In *Fundamentos da metodologia científica* (pp. 320-320).

Li, J., Chen, W. H., Xu, Q., Shah, N., & Mackey, T. (2019, October). Leveraging big data to identify corruption as an SDG goal 16 humanitarian technology. In *2019 IEEE Global*

Humanitarian Technology Conference (GHTC) (pp. 1-4). IEEE.

Martínez-Plumed, F., Casamayor, J. C., Ferri, C., Gómez, J. A., & Vendrell Vidal, E. (2018, September). SALER: a data science solution to detect and prevent corruption in public administration. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 103-117). Cham: Springer International Publishing.

Ministério Público do Estado de Sergipe. (2011). "Portaria n.º 635/2011". Institui o Grupo de Atuação Especial de Combate às Organizações Criminosas – GAECO. *Diário Oficial do MPSE*. Aracaju, Sergipe, Brasil.

Modrusan, N., Rabuzin, K., & Mrsic, L. (2020). Improving Public Sector Efficiency using Advanced Text Mining in the Procurement Process. In *DATA* (pp. 200-206).

Niessen, M. E. K., Paciello, J. M., & Fernandez, J. I. P. (2020, April). Anomaly Detection in Public Procurements using the Open Contracting Data Standard. In *2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG)* (pp. 127-134). IEEE.

Padhi, S. S., & Mohapatra, P. K. (2011). Detection of collusion in government procurement auctions. *Journal of Purchasing and Supply Management*, *17*(4), 207-221.

Popa, M. (2019). Uncovering the structure of public procurement transactions. *Business and Politics*, *21*(3), 351-384.

Pramanik, A., Sarker, A., Islam, Z., & Hashem, M. M. A. (2020, December). Public Sector Corruption Analysis with Modified K-means Algorithm Using Perception Data. In *2020 11th International Conference on Electrical and Computer Engineering (ICECE)* (pp. 198-201). IEEE.

Rabuzin, K., & Modrusan, N. (2019). Prediction of Public Procurement Corruption Indices using Machine Learning Methods. In *KMIS* (pp. 333-340).

Ralha, C. G., & Silva, C. V. S. (2012). A multi-agent data mining system for cartel detection in Brazilian government procurement. *Expert Systems with Applications*, *39*(14), 11642-11656.

Santos, C. M. D. C., Pimenta, C. A. D. M., & Nobre, M. R. C. (2007). A estratégia PICO para a construção da pergunta de pesquisa e busca de evidências. *Revista latino-americana de enfermagem*, *15*, 508-511.

Silva, Luiz Felipe Carvalho. (2022). O Uso de Red Flags no Enfrentamento à Corrupção: *Data Analytics* e Tutela do Patrimônio Público. *Ed Juspodivm.* ISBN: 9786556806518.

Simon, R., & Aalbers, G. (2019). The capacity to combat corruption (CCC) index. *AS/COA*, 15.

Souza, L. B., Júnior, M. C., Silva, R., Fontes, R., Silva, C., Paiva, J., ... & Lins, G. (2023). Data science to identify crimes against public administration.

Tripp, D. (2005). Pesquisa-ação: uma introdução metodológica. *Educação e pesquisa*, *31*, 443-466.

Velasco, R. B., Carpanese, I., Interian, R., Paulo Neto, O. C., & Ribeiro, C. C. (2021). A decision support system for fraud detection in public procurement. *International Transactions in Operational Research*, *28*(1), 27-47.

APÊNDICE

Formulário de Avaliação do Galactus Snitch

Formulário para avaliação da aplicação Galactus Snitch - uma aplicação para apontamento de possíveis irregularidades em contratos públicos

SQ1 - Na busca ativa por fraudes em licitações, a aplicação é capaz de reduzir em mais de 90% o escopo de busca por irregularidades, apontando contratos potencialmente fraudulentos?

a) Concordo totalmente
b) Concordo parcialmente
c) Neutro
d) Discordo parcialmente
e) Discordo totalmente

SQ2 - Com base nas informações exibidas pela aplicação, já é possível selecionar contratos e licitações para apreciação, a partir dos quais poderiam futuramente resultar na abertura de processo de investigação criminal?

a) Concordo totalmente
b) Concordo parcialmente
c) Neutro
d) Discordo parcialmente
e) Discordo totalmente

SQ3 - As redflags utilizadas na aplicação possuem alta relevância na tarefa de identificar fraudes em contratos e licitações públicas?

a) Concordo totalmente
b) Concordo parcialmente
c) Neutro
d) Discordo parcialmente
e) Discordo totalmente

SQ4 - As redflags utilizadas na aplicação são suficientes na tarefa de identificar fraudes em contratos e licitações públicas?

a) Concordo totalmente
b) Concordo parcialmente
c) Neutro
d) Discordo parcialmente
e) Discordo totalmente

SQ5 - Em caso de discordância na questão anterior, aponte possíveis redflags que poderiam ser usadas para aprimorar a ferramenta (opcional)