

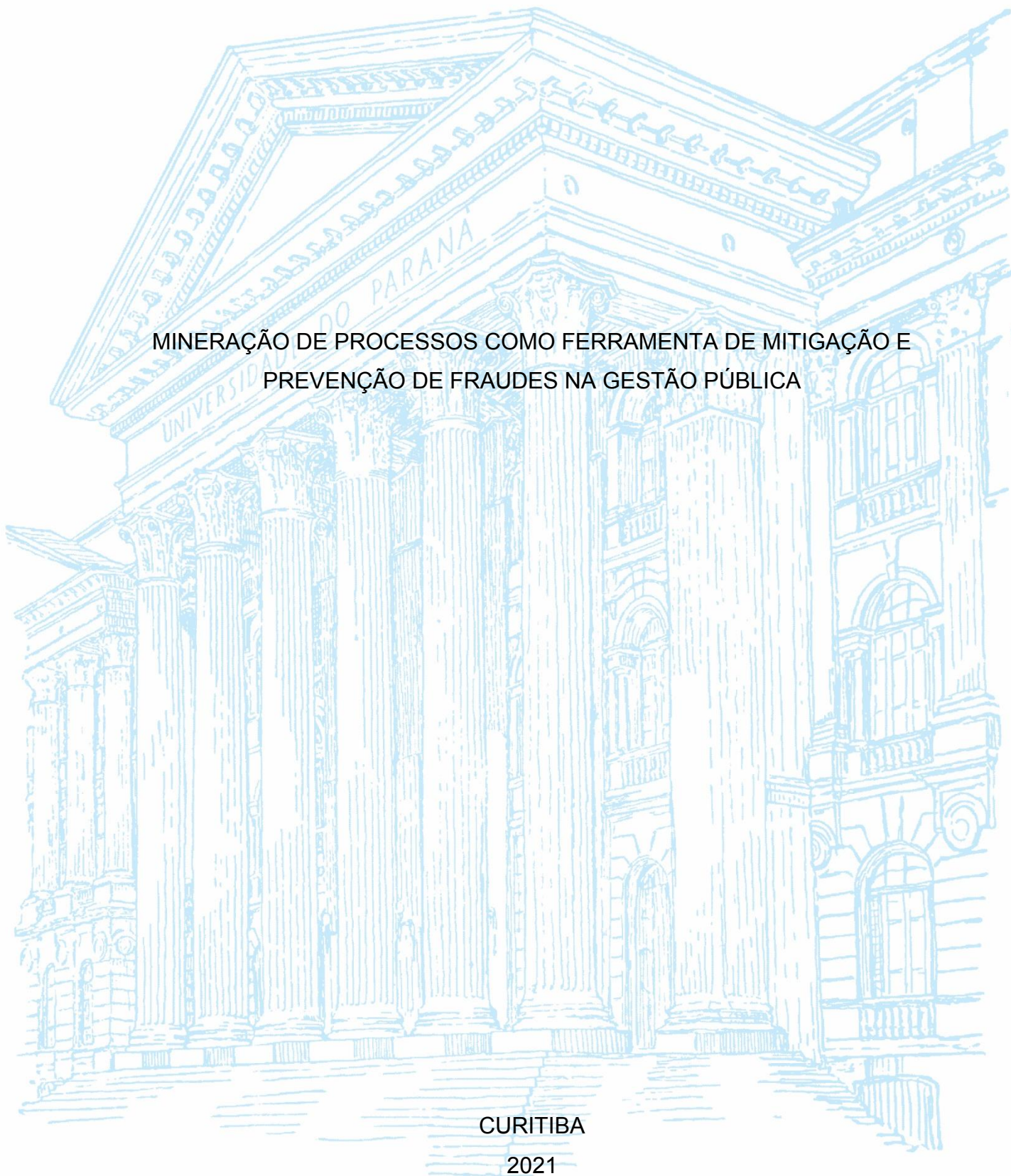
UNIVERSIDADE FEDERAL DO PARANÁ

RICARDO SHIGUERU FUJIWARA

MINERAÇÃO DE PROCESSOS COMO FERRAMENTA DE MITIGAÇÃO E
PREVENÇÃO DE FRAUDES NA GESTÃO PÚBLICA

CURITIBA

2021



RICARDO SHIGUERU FUJIWARA

MINERAÇÃO DE PROCESSOS COMO FERRAMENTA DE MITIGAÇÃO E
PREVENÇÃO DE FRAUDES NA GESTÃO PÚBLICA

Dissertação apresentada ao curso de Pós-Graduação em Gestão da Informação, Setor de Ciências Sociais e Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Gestão da Informação, Área de Concentração: Gestão da Informação e do Conhecimento.

Orientadora: Prof.^a Dr.^a Denise Fukumi Tsunoda

CURITIBA

2021

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DE CIÊNCIAS SOCIAIS
APLICADAS – SIBI/UFPR COM DADOS FORNECIDOS PELO(A) AUTOR(A)
Bibliotecário: Eduardo Silveira – CRB 9/1921

Fujiwara, Ricardo Shigueru

Mineração de processos como ferramenta de mitigação e prevenção
de fraudes na gestão pública / Ricardo Shigueru Fujiwara.- 2021.
98 p.

Dissertação (Mestrado) - Universidade Federal do Paraná. Programa
de Pós-Graduação em Gestão da Informação, do Setor de Ciências
Sociais Aplicadas.

Orientadora: Denise Fukumi Tsunoda.

Defesa: Curitiba, 2021.

1. Gestão da Informação. 2. Mineração de dados. 3. Fraude.
4. Gestão Pública. I. Universidade Federal do Paraná. Setor de Ciências
Sociais Aplicadas. Programa de Pós-Graduação em Gestão da
Informação. II. Tsunoda, Denise Fukumi. III. Título.

CDD 658.4038



TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO DA INFORMAÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **RICARDO SHIGUERU FUJIWARA** intitulada: **MINERAÇÃO DE PROCESSOS COMO FERRAMENTA DE MITIGAÇÃO E PREVENÇÃO DE FRAUDES NA GESTÃO PÚBLICA**, sob orientação da Profa. Dra. DENISE FUKUMI TSUNODA, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 13 de Maio de 2021.

Assinatura Eletrônica

13/05/2021 17:18:04.0

DENISE FUKUMI TSUNODA

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

13/05/2021 17:00:14.0

TAIANE RITTA COELHO

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

16/05/2021 16:42:15.0

ALEXANDRE REIS GRAEML

Avaliador Externo (UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ)

Dedico à minha esposa, companheira e amiga
Mariana Espíndola de Souza pela incentivo e
compreensão durante todos os momentos.

AGRADECIMENTOS

Por mais uma etapa concluída agradeço as pessoas que diretamente ou indiretamente contribuíram nessa trajetória, que assim passo a listar, não em ordem de importância, mas na sequência da minha lembrança.

Novamente agradeço a minha esposa Mariana, que me incentivou a terminar este trabalho. Aos meus pais, Reinaldo e Alice pelo empenho e dedicação na minha formação desde o meu nascimento, e assim se vão alguns de anos de educação. E minhas lindas e queridas sobrinhas Letícia e Amanda por me divertirem nos fins de semana.

À Universidade Federal do Paraná pela oportunidade, em várias etapas dos estudos. Ao Programa de Pós-Graduação em Gestão da Informação, a Coordenação na figura dos professores Dr. Rodrigo Eduardo Botelho Francisco e Dra. Maria do Carmo Duarte Freitas, as queridas Simone da Silva Batista e Camila Campos Pazoti no auxílio como secretárias do programa.

À minha orientadora Dra. Denise Fukumi Tsunoda minha gratidão. Agradeço a paciência e incentivo durante estes dois anos, pelo conhecimento repassado para que conseguisse terminar este trabalho.

Às professoras Dra. Helena de Fátima Nunes Silva e Dra. Deborah Ribeiro Carvalho, e aos professores Dr. Cícero Aparecido Bezerra e Dr. Ricardo Mendes Junior pelo conhecimento compartilhado nas disciplinas ministradas.

Aos professores Dr. Alexandre Reis Graeml, da UTFPR e Dra. Taiane Ritta Coelho, da UFPR pela participação na banca e pelas auxílio e sugestões.

Aos professores do Me. Paulinho Rene Stefanello, Dr. Cícero José Albano e Dra. Aline Cristina Coletto do IFPR Curitiba, pelo apoio e incentivo no ingresso no mestrado.

Aos meus colegas Lucimara, Aline, Leandra e Victor pelos dias de divagação, conversas e ajuda mútua na finalização dos trabalhos durante o mestrado.

“A principal ideia por trás de sistemas complexos é que o conjunto se comporta de maneiras que não podem ser previstas por seus componentes. As interações importam mais que a natureza das unidades” (TALEB, 2018, p.91)

“Uma pessoa honesta jamais cometerá atos criminosos, mas um criminoso se envolverá prontamente em atos legais” (TALEB, 2018, p.93)

RESUMO

A pesquisa analisa o emprego da mineração de processos como ferramenta de prevenção e mitigação de fraudes na gestão pública, contribuindo na produção de medidas de combate à fraude, minimização de riscos e melhoria na aplicação dos recursos e serviços públicos. Adota-se a abordagem qualitativa, de natureza aplicada. Quanto aos objetivos a pesquisa é descritiva com a utilização de meios bibliográficos e *ex post facto*. As etapas propostas visam apontar as ferramentas, bases de dados, métodos e técnicas de processos de mineração para prevenção de fraudes, por meio de uma revisão da literatura, seguida de um mapeamento comparativo de ferramentas de mineração de processos e uso das técnicas de mineração de processos em um conjunto de registros de eventos criados com base em dados abertos de execução de obras públicas. Na revisão de literatura a estratégia de pesquisa considerou o conjunto de termos mineração de processos e fraudes, que indicaram a predominância da utilização da ferramenta ProM nos estudos, com aplicação em processos de compras e nos setores bancário e de saúde. O mapeamento comparou 8 ferramentas, de códigos-fonte abertos e fechados, indicando características relacionadas aos tipos de arquivos importáveis e exportáveis, além de recursos vinculados aos processos, como descoberta, verificação de conformidade e aprimoramento, dentre outros. As simulações utilizando-se três ferramentas permitiram a identificação de anomalias, como atividades de medição da execução de obras, o que proporciona oportunidades na continuidade dos estudos na abordagem de identificar a visão do usuário da mineração de processos.

Palavras-chave: Mineração de processos. Registros de eventos. Fraude. Anomalia. Administração Pública. Obras Públicas. Gestão da Informação.

ABSTRACT

The research analyzes the use of process mining as a tool for preventing and mitigating fraud in public management, contributing to the production of measures to combat fraud, minimize risks and improve the application of resources and public services. A qualitative approach, of an applied nature, is adopted. As for the objectives, the research is descriptive with the use of bibliographic and ex post facto means. The proposed steps aim to point out the tools, databases, methods, and techniques of mining processes for fraud prevention, through a literature review, followed by a comparative mapping of process mining tools and the use of mining techniques. processes in a set of event records created based on open public works execution data. In the literature review, the research strategy considered the set of terms mining processes and fraud, which indicated the predominance of the use of the ProM tool in the studies, with application in purchasing processes and in the banking and health sectors. The mapping compared 8 tools, from open and closed sources, indicating characteristics related to the types of importable and exportable files, as well as resources linked to processes, such as discovery, verification of compliance and improvement, among others. The simulations using three tools allowed the identification of anomalies, such as activities to measure the execution of works, which provide opportunities for continuing studies in the approach to identifying the user's view of process mining.

Keywords: Process mining. Event logs. Fraud. Anomaly. Public administration. Public works. Information management.

LISTA DE FIGURAS

FIGURA 1 – OS 50 TERMOS MAIS UTILIZADOS NAS PALAVRAS-CHAVE DAS DISSERTAÇÕES DO PPGGI ENTRE 2009 E 2020	23
FIGURA 2 – DELIMITAÇÃO SETORIAL DA PESQUISA.....	25
FIGURA 3 – ESTRUTURA DO PROJETO DE PESQUISA.....	26
FIGURA 4 – REPRESENTAÇÃO DO RELACIONAMENTO DA MINERAÇÃO DE PROCESSOS	27
FIGURA 5 – EXEMPLO DE UMA REDE DE PETRI	33
FIGURA 6 – EXEMPLO DE BPMN	35
FIGURA 7 – CARACTERIZAÇÃO DA PESQUISA.....	45
FIGURA 8 – ROTEIRO DA REVISÃO DA LITERATURA	46
FIGURA 9 – QUESTÕES A SEREM RESPONDIDAS NA REVISÃO DA LITERATURA.....	47
FIGURA 10 – SEQUÊNCIA DE CRIAÇÃO DO PROTOCOLO E MAPEAMENTO....	49
FIGURA 11 – CARACTERÍSTICAS IDENTIFICADOS NAS SOLUÇÕES DE MINERAÇÃO DE PROCESSOS.....	50
FIGURA 12 – EXTRAÇÃO E TRATAMENTO DOS DADOS PARA ANÁLISE	53
FIGURA 13 – ETAPAS PARA A CONSTRUÇÃO DA BASE DE DADOS	54
FIGURA 14 – SEQUÊNCIA DOS TRABALHOS DA MINERAÇÃO DE PROCESSOS	55
FIGURA 15 – GESTÃO DA INFORMAÇÃO, AUDITORIA E GESTÃO DE RISCOS NA IDENTIFICAÇÃO DE FRAUDADORES	56
FIGURA 16 – IMPORTAÇÃO DO ARQUIVO CSV NA FERRAMENTA DISCO.....	68
FIGURA 17 – RETRATO DO CONJUNTO DE REGISTROS.....	68
FIGURA 18 – TELA DE ANIMAÇÃO DO FLUXO DOS EVENTOS DA FERRAMENTA DISCO.....	69
FIGURA 19 – RECORTE DO TEMPO MEDIANO DAS ATIVIDADES	70
FIGURA 20 – SEQUÊNCIA ERRADA.....	70
FIGURA 21 – MODELO DE PROCESSOS NO CELONIS SNAP	71
FIGURA 22 – ANIMAÇÃO DO MODELO DE PROCESSOS NO CELONIS SNAP...	72
FIGURA 23 – MODELO DE PROCESSO NA FERRAMENTA PROM.....	73

LISTA DE QUADROS

QUADRO 1 – AFIRMAÇÕES E LACUNAS COMO INDAGAÇÕES DE PESQUISA.	19
QUADRO 2 – PROTOCOLO DE REVISÃO	47
QUADRO 3 – TRABALHOS RELACIONADOS A PESQUISA	48
QUADRO 4 – FERRAMENTAS MAIS UTILIZADAS NOS ESTUDOS DA REVISÃO SISTEMÁTICA DA LITERATURA	57
QUADRO 5 – BASE DE DADOS UTILIZADAS NA MINERAÇÃO DE PROCESSOS	58
QUADRO 6 – RESULTADO DA REVISÃO DA LITERATURA.....	59
QUADRO 7 – ESPECIFICAÇÃO DA FERRAMENTA DE MINERAÇÃO DE PROCESSOS	63
QUADRO 8 – CARACTERÍSTICAS DE IMPORTAÇÃO E EXPORTAÇÃO DOS REGISTROS DE EVENTOS	64
QUADRO 9 – SUPORTE AS TÉCNICAS DE MINERAÇÃO DE PROCESSOS	65

LISTA DE TABELAS

TABELA 1 – RESULTADO DAS CONSULTAS NOS REPOSITÓRIOS.....	21
TABELA 2 – PESQUISA NAS BASES DE DADOS	22
TABELA 3 – CINCO ARTIGOS MAIS REFERENCIADOS.....	60

LISTA DE ABREVIATURAS OU SIGLAS

BI	- <i>Business Intelligence</i>
BPMN	- <i>Business Process Model and Notation</i>
BPMS	- <i>Business Process Management System</i>
CAPES	- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
EFS	- Entidades Fiscalizadoras Superiores
EPC	- <i>Engineering, Procurement and Construction</i>
ERP	- <i>Enterprise Resource Planning</i>
IBICT	- Instituto Brasileiro de Informação em Ciência e Tecnologia
INCOSAI	- <i>International Congress of Supreme Audit Institutions</i>
INTOSAI	- <i>International Organization of Supreme Audit Institutions</i>
IRB	- Instituto Rui Barbosa
ISLA2020	- 4ª Conferência sobre Sistemas de Informação na América Latina
ML	- <i>Machine Learning</i> / Aprendizado de Máquina
PM	- <i>Process Mining</i> / Mineração de Processos
PPGGI	- Programa de Pós-graduação em Gestão da Informação
SNGIC	- II Seminário Nacional de Gestão da Informação e do Conhecimento
TCE-PR	- Tribunal de Contas do Estado do Paraná
TCU	- Tribunal de Contas da União
UEL	- Universidade Estadual de Londrina
UFSCar	- Universidade Federal de São Carlos
UNICAMP	- Universidade Estadual de Campinas

SUMÁRIO

1 INTRODUÇÃO	16
1.1 ABORDAGEM DA PESQUISA	18
1.2 OBJETIVOS	20
1.3 JUSTIFICATIVA	20
1.4 DELIMITAÇÃO DA PESQUISA	24
1.5 ESTRUTURA DO TRABALHO	25
2 REVISÃO DE LITERATURA	27
2.1 MINERAÇÃO DE PROCESSOS	27
2.1.1 Representação dos processos	31
2.1.1.1 Redes de Petri	32
2.1.1.2 BPMN	34
2.2 FRAUDE	35
2.2.1 Motivação da fraude	36
2.2.2 <i>Outliers</i> ou anomalias em processos	38
2.2.3 Gestão da informação e a fraude	42
3 ENCAMINHAMENTOS METODOLÓGICOS	44
3.1 REVISÃO SISTEMÁTICA DA LITERATURA	45
3.2 PROTOCOLO PARA COMPARAÇÃO DAS FERRAMENTAS	48
3.3 SIMULAÇÃO DO EMPREGO DA MINERAÇÃO DE PROCESSOS	52
3.3.1 Tratamento da base de dados	52
3.3.2 Análise dos resultados da utilização de ferramentas de mineração de processos	55
4 ANÁLISE E DISCUSSÃO DOS RESULTADOS	56
4.1 APONTAMENTOS DA REVISÃO DA LITERATURA	57
4.2 COMPARATIVO ENTRE FERRAMENTAS	62
4.3 SIMULAÇÕES NA FERRAMENTA DE MINERAÇÃO DE PROCESSOS	66
5 CONSIDERAÇÕES FINAIS	75
5.1 REVISÃO SISTEMÁTICA DA LITERATURA	76
5.2 COMPARATIVO DAS FERRAMENTAS DE MINERAÇÃO DE PROCESSOS ...	77
5.3 O EMPREGO DA MINERAÇÃO DE PROCESSOS EM UM CONJUNTO DE REGISTROS DE LICITAÇÕES DE OBRAS PÚBLICAS DE MUNICÍPIOS PARANAENSES.	77

5.4 SUGESTÕES PARA CONTINUIDADE DO ESTUDO	78
REFERÊNCIAS.....	80
APÊNDICE A – DETECÇÃO DE FRAUDES COM APRENDIZADO DE MÁQUINA	88
APÊNDICE B – VISUALIZAÇÃO DE ATIVIDADES NO DISCO (SIMPLIFICAÇÃO DE ATIVIDADES)	90
APÊNDICE C – TELA DA FERRAMENTA PROM	91
ANEXO I – ELABORAÇÃO DOS DIRECIONAMENTOS FUTUROS DA AUDITORIA (FISCALIZAÇÃO) PÚBLICA DECLARAÇÃO	92
ANEXO II – RELACIONAMENTO E DICIONÁRIO DE DADOS DO CONJUNTO OBRAS MUNICIPAIS	96

1 INTRODUÇÃO

A busca de melhorias na administração pública para atender aos anseios e cobranças da sociedade, em especial na execução serviços públicos é um desafio a ser enfrentado pelos gestores, partindo deste viés este trabalho aborda o tema de fraudes na gestão pública. O modelo do Estado de bem-estar social torna os entes públicos provedores de serviços públicos, essenciais e necessários à população, através da implementação de políticas públicas (ARAÚJO, ZULLO, TORRES, 2020).

Serviços estes que precisam ser prestados com qualidade e eficiência. A eficiência é um dos princípios previstos no artigo 37 da Constituição Federal do Brasil, aplicável à administração pública no âmbito da União, dos Estados, do Distrito Federal e dos Municípios. Na “adoção de novos mecanismos, notadamente o uso de inteligência artificial [...] tem o potencial de aumentar exponencialmente o nível de eficiência na prestação de serviços públicos” (ARAÚJO, ZULLO, TORRES, 2020), impactando a configuração de algumas funções estatais e o redimensionamento do aparato estatal.

A Declaração de Moscou apresentou os resultados dos debatidos ocorridos no XXIII Congresso Internacional das Entidades Fiscalizadoras (INCOSAI), proporcionando um direcionamento futuro para as auditorias no setor público, que incluem a necessidade de uma resposta mais eficaz às oportunidades decorrentes dos avanços tecnológicos e o foco na prestação de contas (*accountability*) e na transparência na gestão pública (IRB, 2019). Neste contexto, as Entidades Fiscalizadoras Superiores (EFS) são orientadas:

- As EFS poderiam promover a cultura da disponibilização e abertura dos dados, dos códigos fontes e dos algoritmos;
- As EFS poderiam objetivar um melhor uso de análise de dados em auditorias, incluindo estratégias de adaptação como o planejamento para tais auditorias, o desenvolvimento de equipes experientes em análise de dados, e a introdução de novas técnicas na prática de auditoria (fiscalização) pública;
- As EFS podem fomentar uma mentalidade experimental par reforçar a inovação e o desenvolvimento;
- As EFS são encorajadas a formar os auditores do futuro capazes de: trabalho com análise de dados, ferramentas de inteligência artificial e avançados métodos de análise quantitativa [...] (IRB¹, 2019, p.3-4)

¹ Esta é uma versão traduzida da Declaração de Moscou, elaborada por uma equipe do Instituto Rui Barbosa (IRB) e do Tribunal de Contas do Estado Paraná (TCE-PR). A versão em inglês está disponível em: <https://incosai2019.ru/en/documents/51?download=293>.

Algumas iniciativas já estão em andamento como os desenvolvidos pela Tribunal de Contas da União - TCU, como o SOFIA – Sistema de Orientação sobre Fatos e Indícios para o Auditor, o robô MONICA - Monitoramento Integrado para o Controle de Aquisições (IRB, 2020; ARAÚJO, ZULLO, TORRES, 2020) e o sistema ALICE – Análise de Licitações e Editais que “usa técnicas de mineração de textos para identificar irregularidades em editais de licitação e de pregões eletrônicos publicados no portal eletrônico de compras do Governo Federal” (EBC, 2019).

A fraude e a corrupção existem há séculos e se renovam com as novas tecnologias, porém os métodos e motivos básicos permanecem os mesmos ao longo do tempo. A prevenção tem mais chances de sucesso no combate à fraude, comparando-se com o tempo de recuperação e mesmo a dificuldade de detectar e reaver o dinheiro posteriormente (IYER; SAMOCIUK, 2006). O impacto financeiro causado tanto pelo crime em si, quanto pelas medidas de prevenção da fraude, vem aumentando durante os anos, como identificou a pesquisa da empresa de auditoria PwC, nas empresas brasileiras que foram vítimas de fraude e crimes econômicos. Essas fraudes atingiam 12% das empresas consultadas em 2016 passando para 50% das empresas em 2018, que acarretou a necessidade de aumentar as despesas no combate à fraude (PWC, 2018).

Mudanças operacionais são necessárias para a aplicação de sistemas de segurança para prevenção e detecção fraudes, que em meio a transformação digital provocam o aparecimento de novas tecnologias. Conforme Van der Aalst (2016) a mineração de processos permite uma verificação de conformidade de forma mais rigorosa, além de promover uma validação e averiguação da confiabilidade das informações sobre os processos organizacionais, com isso revelando desvios e defeitos, sendo uma valiosa ferramenta para análise de conformidade e gestão de riscos.

A necessidade de implementação de métodos que auxiliem auditores, gestores e órgãos fiscalizadores a analisarem um volume de dados de maneira mais eficiente torna-se essencial. As premissas que orientam as entidades fiscalizadoras no âmbito mundial, já vislumbram ações pelos tribunais de contas e órgãos de controle no Brasil buscando novas ferramentas que auxiliem no controle e monitoramento das ações governamentais. Assim, esta pesquisa busca apresentar e analisar a mineração de processos como ferramenta auxiliar aos analistas e gestores públicos.

1.1 ABORDAGEM DA PESQUISA

As pesquisas apontam vários direcionamentos ferramentas, técnicas ou métodos como a utilização de *blockchain*, assinatura digital e criptografia em processos de aquisição (THIO-AC et al., 2019), o *blockchain* como tecnologia de proteção da identidade em contratos bancários, e como requisito legal devido a imposição de regulamentos de proteção de dados já existente em várias partes do mundo (WARDHANA; DANTES; AYANTO; 2020) e sistemas de *business intelligence* com aplicação conjunta de métodos estatísticos (PILON et al., 2015).

Ainda aplicações da mineração de dados na identificação de fraude em operações de cartão crédito (BHATTACHARYYA et al., 2011), do aprendizado de máquina na identificação de contas falsas em redes sociais (XIAO; FREEMAN; HWA, 2015); predição e mitigação de fraudes com a utilização de técnicas de aprendizado de máquinas em cadeias de suprimentos (CONSTANTE-NICOLALDE; GUERRA-TERÁN; PÉREZ-MEDINA, 2020) e em seguros-saúde (BAUDER; ROSA; KHOSHGOFTAAR, 2018), e técnicas de mineração de processos para identificação de fraudes em processos de negócios (VAN DER AALST; DE MEDEIROS, 2005; HUDA et al., 2014; FEBRIYANT; SARNO; EFFENDI, 2018; CAESARITA; SARNO; SUNGKONO, 2018).

A mineração de processos tem a capacidade de extrair, de forma não tendenciosa, modelos a partir de um conjunto de registros transacionais, o que permite encontrar falhas ou identificar fraudes nos processos investigados, auxiliando no monitoramento de controles internos (JANS et al., 2011). Porém, a relação entre o grande número transações normais comparada com o menor número de transações fraudulentas é um ponto que dificulta a modelagem de algoritmos de prevenção e identificação de fraudes (WANG; LI; ZHAO, 2019). Outro ponto a ser considerado, é a condição dinâmica e contínua da geração de dados, que requer análises em tempo real (durante a execução) para identificação de casos anômalos, e não apenas com registros de eventos completos (TAVARES et al., 2018).

A dificuldade de acesso e visualização de algoritmos computacionais pode ocorrer, motivados por questões de: sigilo corporativo, como forma de proteção comercial; desconhecimento técnico, no aspecto de interpretação dos algoritmos computacionais; e pôr fim pela complexidade ou escala de aplicação de plataformas

formada por multicomponentes, que impede operacional do sistema como um todo. (BURRELL, 2016). Outro envolvido na falta de transparência dos algoritmos aplicados nos processos decisões e a sua conciliação com princípio constitucional da publicidade, isto é, com a Lei de Acesso à Informação (ARAÚJO, ZULLO, TORRES, 2020), alinhando-se as recomendações de “abertura dos sistemas informatizados que embasam a tomada de decisões utilizada pelo governo, sob os princípios de códigos-fonte abertos e dados abertos” dispostas na Declaração de Moscou (IRB, 2019).

A complexidade nos modelos, implementação e análise de negócios induzem ao desenvolvimento de novas tecnologias e alterações organizacionais, redefinindo as funções tradicionais do gerenciamento de processos (BEVERUNGEN et al., 2020). Uma mudança de paradigma nas funções operacionais do Estado, com a introdução de inovações tecnológicas, efetividade na prestação dos serviços públicos e proteção dos recursos públicos. A aplicação da mineração da processos para os fins de mudança inovativa poderia ser uma solução na identificação de fraudes, ainda opaco no âmbito no desconhecimento técnico sobre seus benefícios, usabilidade e aplicação em conjuntos de dados desequilibrados, desta forma esta pesquisa tem como abordagem as afirmações e lacunas dispostas no Quadro 1.

QUADRO 1 – AFIRMAÇÕES E LACUNAS COMO INDAGAÇÕES DE PESQUISA

Área	Afirmações e Lacunas
Mineração de Processos	<ul style="list-style-type: none"> • A mineração de processos realiza uma verificação de conformidade de forma mais rigorosa que permite revelar desvios e defeitos (VAN DER AALST, 2016) • A transparência no acesso dos algoritmos é dificultada devido ao desconhecimento técnico e a complexidade (BURRELL, 2016)
Fraudes	<ul style="list-style-type: none"> • A mineração de processos apresenta oportunidades relacionadas a aplicação de alarmes para detecção de anomalias ou detecção de desvios (GARCIA et al., 2019) • A análise de dados desequilibrados, onde a relação entre o número transações normais e transações fraudulentas, dificulta a modelagem de algoritmos de prevenção e identificação de fraudes (WANG; LI; ZHAO, 2019)

FONTE: Elaborado pelo autor (2021).

É neste contexto de identificar fraudes e desvios, procura-se ratificar ou retificar as afirmações apresentadas e preencher as lacunas, com base nesta questão

norteadora da pesquisa: **Como a mineração de processos pode auxiliar na mitigação e prevenção de fraudes na gestão pública?**

1.2 OBJETIVOS

Para responder à questão de pesquisa traçou-se o seguinte objetivo geral: **Avaliar a empregabilidade da mineração de processos como ferramenta de mitigação e prevenção de fraudes.** Para avistar esse objetivo trilharam-se os seguintes objetivos específicos:

- a) apontar ferramentas, bases de dados, métodos e técnicas de mineração de processos empregados na prevenção de fraudes, através de uma revisão da literatura;
- b) comparar ferramentas de mineração de processos, indicando as principais características de usabilidade;
- c) selecionar e julgar o emprego da mineração de processos em um conjunto de registros de licitações de obras públicas de municípios paranaenses.

1.3 JUSTIFICATIVA

A contribuição social e econômica surge da necessidade de produzir medidas que visem a prevenção de atos fraudulentos e a mitigação dos riscos que estão sujeitos as organizações públicas e conseqüentemente a sociedade. Estes riscos abarcam não somente recursos financeiros, mas produzem efeitos que podem acarretar riscos à saúde e a própria vida. O tema “fraude” espalha-se e é observado nos mais variados momentos como nas operações bancárias e financeiras, em compras no comércio eletrônico, em licitações públicas, nas mídias sociais, no envio de notícias, causando danos incalculáveis na sociedade.

No campo acadêmico, primeiramente realizou-se uma busca das contribuições científicas nos repositórios do Catálogo de Teses e Dissertações da CAPES e da Biblioteca Digital Brasileira de Teses e Dissertações do IBICT, com a aplicação do critério de pesquisa o termo “mineração de processos” e o termo em inglês “*process mining*”, que resultou em 39 pesquisas localizadas no repositório da

CAPES e 28 no repositório do IBICT. Do resultado excluíram-se os resultados duplicados, assim restando 49 contribuições científicas. A Tabela 1 apresenta os resultados das consultas realizadas nos bancos de teses e dissertações, detalhadas por ano da defesa, e por instituição, considerando-se somente as publicações disponíveis digitalmente.

A Pontifícia Universidade Católica do Paraná - PUC-PR aparece com 8 contribuições, seguida da Universidade de São Paulo – USP (5) e da Pontifícia Universidade Católica do Rio de Janeiro (5). As duas dissertações produzidas na Universidade Federal do Paraná foram produzidas por pesquisadores vinculados ao Programa de Pós-Graduação em Engenharia da Produção e aplicam a mineração de processos na área da saúde.

O tema mineração de processo vem crescendo nos últimos anos, apontando para o interesse em termos de estudos. Na UFPR o direcionamento em pesquisas na área da saúde, que demonstra um campo ainda ser explorado nas pesquisas e contribui para o desenvolvimento de trabalhos na temática, principalmente na aplicação em outros setores, como é o caso da gestão pública.

TABELA 1 – RESULTADO DAS CONSULTAS NOS REPOSITÓRIOS

Ano da Defesa	Distribuição por Instituição Dissertações ou Teses na CAPES e IBICT
2006	UFRGS (1)
2009	UNIRIO (1)
2011	UFOP (1), UNICAMP (1)
2012	UFRGS (1)
2013	PUC-PR (1), UFPE (2), UFRJ (1), UFRN (1), UNB (1)
2014	UFRGS (1)
2015	PUC-PR (2), UNIRIO (3), USP (1)
2016	FURG (1), UECE (1), UEL (1), UNB (2)
2017	PUC-PR (2), PUC-RIO (2), UFRJ (1), UFRN (1), UNESP (1), UNISINOS (1), USP (2)
2018	PUC-PR (3), PUC-RIO (1), UFES (1), UFG (2), UFPR (1), UFRJ (1), UNB (1)
2019	PUC-RIO (2), UEL (1), UFPR (1), USP (2)

FONTE: Elaborada pelo autor (2020).

Como parte integrante do desenvolvimento desta pesquisa levou-se a divulgação o artigo “Mineração de Processos na Identificação de Fraudes: um Escopo da Produção Científica”, que foi aprovado para apresentação e publicado nos anais da 4ª Conferência sobre Sistemas de Informação na América Latina – ISLA 2020, em 11 de agosto de 2020, indicando o interesse perante o meio acadêmico e que justificou a continuidade da pesquisa.

O artigo consistiu numa pesquisa bibliométrica com buscas nas bases Scopus, Web of Science e Ebsco (*Computer & Applied Sciences Complete, Information Science & Technology Abstracts e na Library, Information Science & Technology Abstracts*) (FUJIWARA; TSUNODA, 2020).

Como critério de busca utilizou-se os termos “mineração de processos” e “fraude”, com os respectivos termos em inglês, resultando em 70 artigos, que excluídos as publicações duplicadas resultou no escopo de 35 artigos. Estes artigos foram publicados em periódicos e anais de conferência, abrangendo 28 fontes de pesquisas, identificando-se a concentração das pesquisas em alguns autores e universidades (FUJIWARA; TSUNODA, 2020). A Tabela 2 contém os resultados da pesquisa, de acordo a estratégia selecionada:

TABELA 2 – PESQUISA NAS BASES DE DADOS

Palavras e/ou Termos	Scopus	Web of Science	Ebsco
“mineração de processos” ou “process mining”	3.407	1.863	488
(“mineração de processos” ou “process mining”) e (“fraude” ou “fraud”)	35	27	8

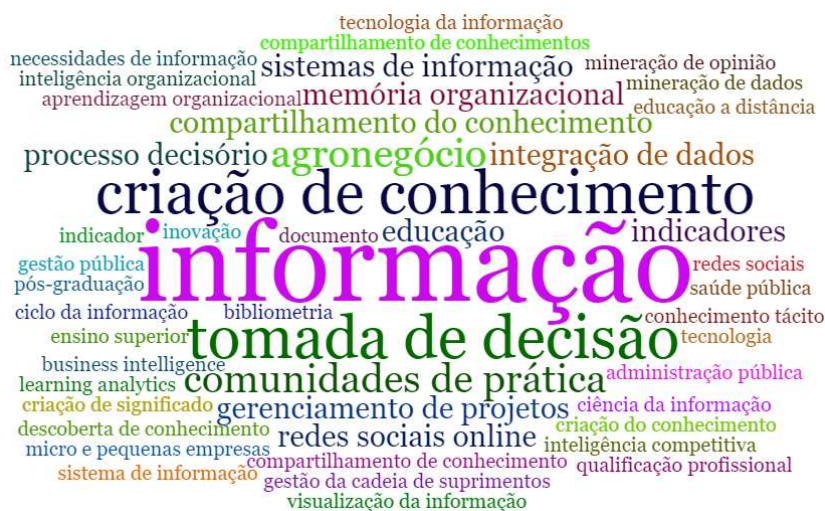
FONTE: Elaborada pelo autor (2020).

O presente estudo adere-se aos objetivos propostos pelo Programa de Pós-Graduação em Gestão da Informação (PPGGI) no que tange a interdisciplinaridade e estudo da informação, quanto a utilização de tecnologias voltadas a gestão da informação. No tocante a interdisciplinaridade, a identificação de fraudes abrange diversas áreas do conhecimento como medicina, engenharia, finanças e no presente estudo a gestão pública.

No aspecto tecnológico, a descoberta de conhecimento através do emprego de ferramentas e algoritmos de mineração de processos produz soluções que abrange etapas como o mapeamento de processo, monitoramento e controle, que neste aspecto alinha-se aos objetivos da linha de pesquisa Informação, Tecnologia e Gestão. A diretriz desta linha de pesquisa é o estudo das informações tendo a tecnologia como fator determinante na efetividade gerencial, além disso, investiga o desenvolvimento de métodos, técnicas e ferramentas que busquem transformar a informação como insumo para o desenvolvimento de processos.

Agrega-se a justifica-se o presente estudo no PPGGI, levantaram-se as palavras-chaves dos autores em 116 dissertações defendidas no programa até o ano de 2020. Identificam-se os termos como: informação, criação do conhecimento, compartilhamento de conhecimento, sistemas de informação, ciclo da informação, ciência da informação e visualização da informação e termos relacionados a diversas áreas de conhecimento, que demonstra a interdisciplinaridade das pesquisas do programa. A Figura 1 apresenta as 50 palavras-chave mais utilizadas nas dissertações do PPGGI defendidas entre os anos de 2009 e 2020.

FIGURA 1 – OS 50 TERMOS MAIS UTILIZADOS NAS PALAVRAS-CHAVE DAS DISSERTAÇÕES DO PPGGI ENTRE 2009 E 2020



FONTE: Elaborada pelo autor (2020).

Entre os termos apresentados na Figura 1 relacionam-se a área de negócios como tomada de decisão, processo decisório, gerenciamento de projetos, micro e pequenas empresas ou agronegócio, na educação, desde educação à distância,

ensino superior ou pós-graduação, setores públicos envolvendo a gestão pública ou a saúde pública, chegando as redes sociais e o uso de tecnologias de informação.

Neste contexto, a pesquisa relacionada às técnicas de mineração de processos, em especial a aplicação desta na identificação de fraudes, adiciona um conhecimento ainda não explorado no PPGI.

No aspecto pessoal, o tema do estudo associa-se à experiência profissional do pesquisador na área de auditoria fiscal-tributária. O trabalho de auditoria envolve um processo de análise, com intuito de identificar possíveis omissões de recolhimento de tributos, e muitas vezes de fraudes. Acrescenta-se o grande volume de dados e informações distribuídos em documentos físicos e digital, e a necessidade da busca de novos conhecimentos e ferramentas que contribuam facilitem a execução mais célere das atividades.

Com as justificativas para o presente estudo, na sequência apresentam-se os seus delimitadores considerando-se a inviabilidade do desenvolvimento da pesquisa ter uma abrangência que incluam as mais diversas variáveis e condicionantes.

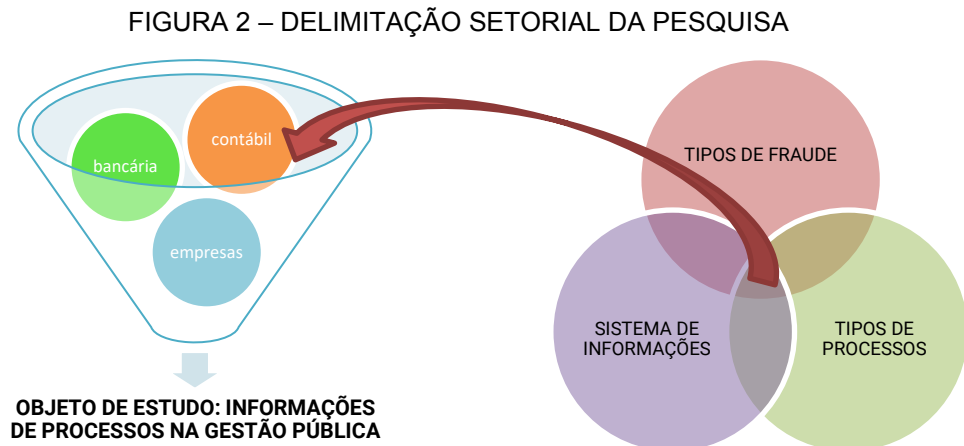
1.4 DELIMITAÇÃO DA PESQUISA

A fraude aparece em diferentes ambientes, momentos e oportunidades, como falsificação de produtos, a criação de empresas falsas, páginas falsas para roubo de informações pessoais, em documentos pessoais, licitações, nos procedimentos tributários, contábeis, e continua a lista. Como delimitação do tema, nesta pesquisa aborda-se a fraude nos processos da gestão pública, onde envolve-se uma sequência cronológica de atividades, ações de tomada de decisão manuais ou automatizados.

A sequência de atividades que caracteriza um processo nem sempre é clara ou está estabelecida, considera-se a possibilidade de existir etapas de processos executados de forma aleatória ou não serem executados. No caso da aplicação da mineração de processos e a busca da identificação de fraudes existe uma necessidade de identificar padrões de processos, que sejam recorrentes e possibilitem uma análise comparativa, sendo um limitador dos tipos de processos a serem abordados nesta pesquisa.

O elemento informação na mineração de processos, relaciona-se a ação de registrar, cada atividade finalizada envia uma “informação” a próxima atividade para

que ela inicie. O marco temporal de início e término da atividade é essencial para a aplicação da mineração de processos. Do ponto de vista da Gestão da Informação, esta pesquisa delimita-se à confluência dos três elementos: fraude, processo e informação, conforme apresentada na Figura 2, e setorizado no âmbito da gestão pública.



FONTE: Elaborada pelo autor (2021).

Do ponto de vista amplo, a delimita-se a pesquisa na técnica de mineração de processos no ambiente de inter-relacionamento destes três elementos, isto é, nas informações de processos executados no âmbito da gestão pública. O desenvolvimento ocorrerá conforme estrutura de trabalho indicada na subseção 1.5.

1.5 ESTRUTURA DO TRABALHO

Inicia-se o trabalho pela introdução, se desenvolvendo de acordo com o escopo descrito na Figura 3, que se divide em cinco seções.

FIGURA 3 – ESTRUTURA DO PROJETO DE PESQUISA



FONTE: Elaborada pelo autor (2020).

Após a introdução, passa-se a próxima seção, que contempla a revisão da literatura, onde se conceitua conceitos necessários ao desenvolvimento e compreensão da pesquisa.

2 REVISÃO DE LITERATURA

O embasamento teórico que orientou a resolução da questão de pesquisa será apresentado nesta seção, primeiramente conceitua-se a técnica escolhida, a mineração de processos e suas características técnicas relacionadas aos modelos relacionados de visualização e na sequência o arcabouço envolvendo a temática fraude, onde se inclui o fator motivador, o ambiente e as situações influenciadoras.

2.1 MINERAÇÃO DE PROCESSOS

Uma disciplina que fica entre a mineração de dados e a inteligência computacional, e a modelagem e análise de processos de negócios, a mineração de processos tem o intuito de extrair conhecimento através da descoberta e monitoramento de registros de eventos disponíveis em sistemas de informação. A mineração de processos, permite a extração de modelos que propiciam melhorias dos processos de negócios, possibilitam a análise de processos através da verificação de conformidades, com isso monitorando possíveis desvios (VAN DER AALST et al., 2012). A Figura 4 apresenta a mineração de processos e alguns relacionamentos.

FIGURA 4 – REPRESENTAÇÃO DO RELACIONAMENTO DA MINERAÇÃO DE PROCESSOS



FONTE: Elaborada pelo autor (2020) com base em Van der Aalst (2012)

O relacionamento da mineração de processos com gestão de negócios, é vislumbrado na busca da melhoria do desempenho operacional, constantes em abordagens como a gestão de qualidade total, melhoria contínua de processos, melhoria dos processos de negócios e de práticas que propõem a eliminação de defeitos nos processos, como o *six sigma* (VAN DER AALST et al., 2012). Técnicas de mineração de processos “oferecem um meio de verificar de forma mais rigorosa a conformidade e determinar a validade e confiabilidade das informações sobre os processos centrais de uma organização” (VAN DER AALST et al., 2012, tradução minha).

O registro de eventos é ponto inicial da mineração de processos, que pressupõe um conjunto de eventos sequenciais, com atividades bem definidas em um processo e relacionadas a uma instância de processo. Além do fluxo de trabalho, a mineração de processos utiliza-se de informações extras como o recurso empregado, ou seja, o responsável pela execução ou dispositivo utilizado, data e hora do evento, e outros dados registros com o evento (VAN DER AALST et al., 2012). As técnicas de mineração de processos segundo sua finalidade podem ser agrupadas em três classes, conforme Munoz-Gama (2016):

- a) descoberta: esta técnica utiliza-se o registro de eventos como entrada e produz um modelo a partir do conhecimento obtido a partir dos próprios registros;
- b) conformidade: utiliza-se além do registro de eventos agrega o modelo de processo como entrada. Tem como objetivo a verificação do comportamento modelado comparando com o comportamento observado nos registros. A conformidade pode verificar além atualização de documentos de processos, permite identificar se o modelo de processo descoberto representa fielmente o processo real;
- c) aprimoramento: Assim como a verificação de conformidade, o aprimoramento também se utiliza como entrada um modelo e registros de eventos, esta comparação visa o melhoramento do modelo de acordo com as informações constantes nos registros. Este aprimoramento, pode ser

executado através de uma reparação do modelo ou da adição de informações adicionais ao modelo.

Outras perspectivas relacionadas a mineração de processos são apresentadas por Burattin (2015), que fogem da simples consideração de descoberta do fluxo de controle, mas trazem interessantes informações:

- a) perspectiva organizacional: a mineração social que consiste na realização de uma análise da rede social com base nos dados obtidos nos processos de negócios. Nesta perspectiva podemos ter duas abordagens: a primeira consiste na análise da rede de conexões como um todo, desta forma considerando um conjunto completo de interações existentes dentro de um grupo de pessoas, e a segunda abordagem concentra análise a partir de um indivíduo específico e o conjunto de interações com outras pessoas;
- b) perspectiva da verificação de conformidade: tem o intuito de comparar o comportamento observado nos registros de eventos, com o esperado, disposto no modelo, assim possibilitando a descoberta de inconsistências. Esta verificação pode ser constatada em duas atividades: no alinhamento do negócio, avalia-se através da comparação entre registros e modelo, e na auditoria, o que é avaliado se o processo atual atende aos limites impostos pelos gestores ou por leis;
- c) perspectiva dos dados: esta perspectiva refere-se à consideração de integrar de outros tipos de dados aos existentes nos fluxos de controle.

Além disso, temos algumas características da mineração de processos apontadas por Van der Aalst et al. (2012), e que refutam alguns equívocos relacionados a aplicação desta técnica.

- a) mineração de processos não se limita à descoberta de modelos de processos, em registros de eventos, sendo essa apenas uma das funções que agregam a descoberta, a verificação de conformidade e ao aprimoramento, melhoria no processo. Outro ponto a salientar, é que o escopo de trabalho também não se limita a análise do fluxo de trabalho,

sendo que as perspectivas de funções e tempo também possuem papel importante;

- b) mineração de processos não é apenas um tipo específico de mineração de dados, pois as maiores das técnicas de mineração de dados não são centradas no processo, sendo necessários novas representações e algoritmos;
- c) a mineração de processos não se limita à análise *offline*. Apesar da aplicação das técnicas na extração de dados de eventos históricos, também é possível a aplicação em casos em execução.

No caso da verificação de conformidade em mineração de processos, o conceito desta técnica pode ter abordagens diferenciadas. As diferentes formas conceituais são listadas no estudo de Elhagaly et al. (2019):

- encontro de inconsistências entre um processo executados e o comportamento modelado;
- concentra em estabelecer até que ponto as execuções reais de um processo estão alinhadas com o comportamento esperado de um modelo referencial;
- encontro de inconsistências entre um registro de eventos e o comportamento modelado e quantificado em métricas;
- alinhamento do comportamento modelado e o observado;
- verificação da ordem de execução;
- encontro de inconsistências entre um registro de eventos e o comportamento modelado e a predição de infrações futuras.

Como importante ferramenta para as organizações, a mineração de processos tem sua utilidade em processos operacionais não triviais, e apresentam alguns desafios importantes relatados por Van der Aalst et al. (2012):

- a) encontrar, mesclar e limpar dados de eventos, com o objetivo de extrair dados adequados para a mineração de processos, dificultado pela origem dos dados de diversas fontes, os dados dos eventos estarem geralmente

centrados no objeto e não no processo, além da existência de dados incompletos, com ruído, ou referirem-se a diferentes níveis de granularidades ou contextos;

- b) lidar com registros de eventos extremamente grandes ou no extremamente pequenos, além de registros de eventos com diversos tipos de características;
- c) a criação de *benchmarks* representativos, considerando a diversidade de funcionalidade e desempenho dos produtos para aplicação em mineração de processos, além de diferentes técnicas de descoberta de processos, que provoca necessidade da criação de métricas de avaliação dos resultados obtidos na mineração de processos;
- d) lidar com as alterações do processo enquanto é analisado, que podem ocorrer devido as mudanças periódicas ou sazonais, nas condições ambientais, como alterações de mercado.

Outros desafios são a melhoria da visualização da representação dos modelos descobertos, a aplicação em cadeias de suprimentos e em computação em nuvem, o fornecimento de suporte operacional para análises quando elas ocorrem, e a combinação da mineração de processos com outros tipos de análise ou técnicas, o melhoramento da usabilidade e a compreensibilidade para que não-especialistas (VAN DER AALST et al., 2012).

Citado como desafio, a visualização é um aspecto importante a ser apresentado, na sequência serão apresentadas algumas formas de representação dos processos.

2.1.1 Representação dos processos

Os processos de negócios devem ser contemplados por alguns requisitos, salientados por Burattin (2015):

- a) a existência de um conjunto finito de atividades e se suas execuções estão parcialmente ordenadas;

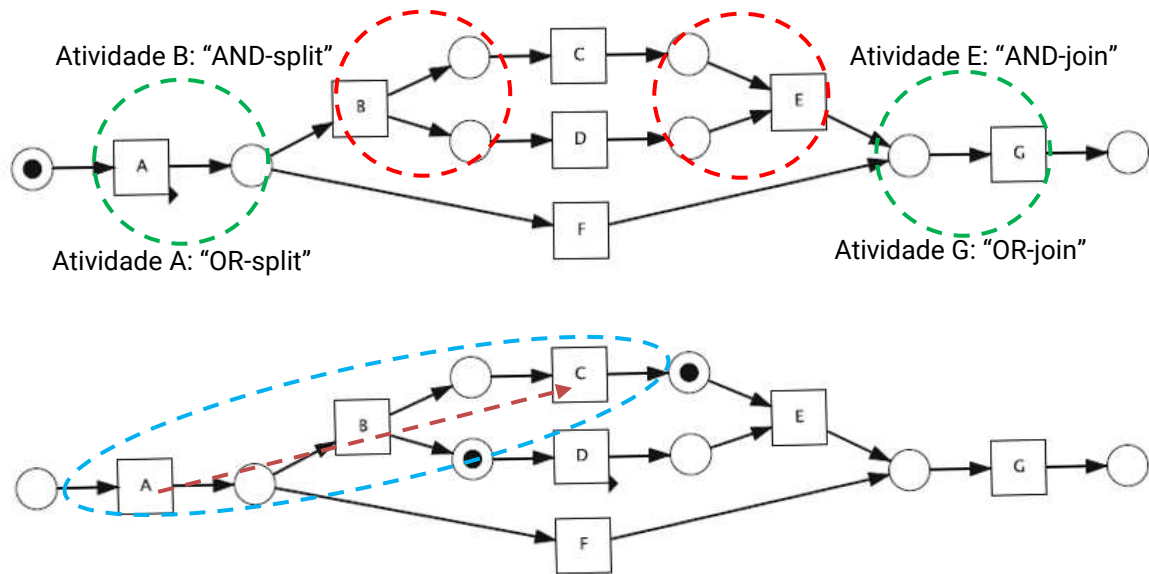
- b) verificação se cada atividade é executada por um ou mais recursos, que podem ser humanos e/ou equipamentos;
- c) se cada atividade executada produz alguma saída que pode ser usada pelas atividades seguintes.

Modelos de negócios claros possuem benefícios como: a possibilidade de aumentar a visibilidade das atividades, que permite a identificação de problemas e de áreas potenciais para otimização e melhoria; o segundo benefício é o agrupamento de acordo com o desenho que melhor permite a definição das atribuições e atividades de auditoria e avaliação. Muitas linguagens permitem a modelagem dos processos negócios, que geralmente são baseadas em grafos, onde os nós são as tarefas e os arcos representam a relação ordenada das tarefas (BURATTIN, 2015). As duas formas de representação baseadas em grafos mais importantes são a redes de Petri e o BPMN (*Business Process Model and Notation* ou Modelo de Processo de Negócios e Notação).

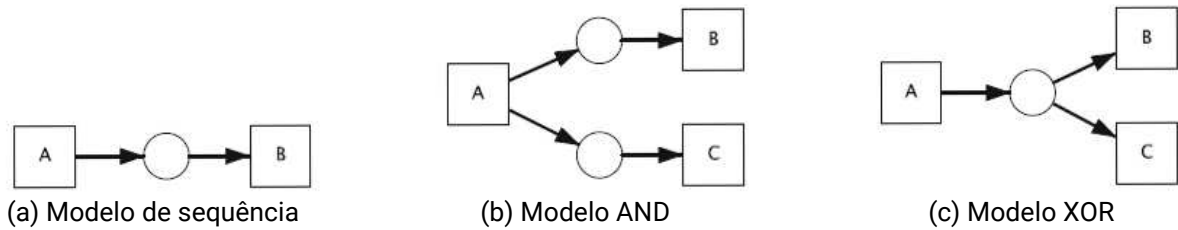
2.1.1.1 Redes de Petri

Proposta por Carl Adam Petri, em 1962, as redes de Petri constituem uma linguagem gráfica que objetiva a representação de um processo, composto por um grafo bipartido onde dois tipos de nós podem estar presentes, definindo transições, que representam atividades que podem ser executadas e locais, que representam os estados que o processo pode atingir, podendo ser intermediários ou finais (BURATTIN, 2015). A Figura 5 apresenta um exemplo de Rede de Petri.

FIGURA 5 – EXEMPLO DE UMA REDE DE PETRI



Situação após a execução das atividades A, B e C. A única transição habilitada, nesta fase, é D.



FONTE: Traduzida de Burattin (2015)

A redes de Petri são estudas por vários pontos de vista, conforme Burattin (2015):

Uma rede de Petri é uma tupla (P, T, F) onde: P é um conjunto finito de lugares; T é um conjunto finito de transições, tal que $P \cap T = \emptyset$, e $F \subset (P \times T) \cup (T \times P)$ é um conjunto de arcos direcionados, chamados de relação de fluxo. O modelo de sequência descreve a dependência causal entre duas atividades (a atividade B requer a execução de A); o modelo AND representa a ramificação simultânea de dois ou mais fluxos (uma vez que A é encerrado, B e C podem iniciar, sem ordem específica e simultaneamente); o modelo XOR define a exclusão mútua de dois ou mais fluxos (uma vez que A é finalizado, apenas B ou C podem iniciar). (BURATTIN, 2015, tradução minha).

A semântica dinâmica de uma rede de Petri é baseada na regra de disparo (ordem de execução), uma transição pode disparar se todos os seus locais de entrada contiverem pelo menos um sinal. O disparo de uma transição gera um sinal para todos

os seus locais e a marcação refere-se a distribuição dos sinais entre os locais de uma rede (BURATTIN, 2015).

2.1.1.2 BPMN

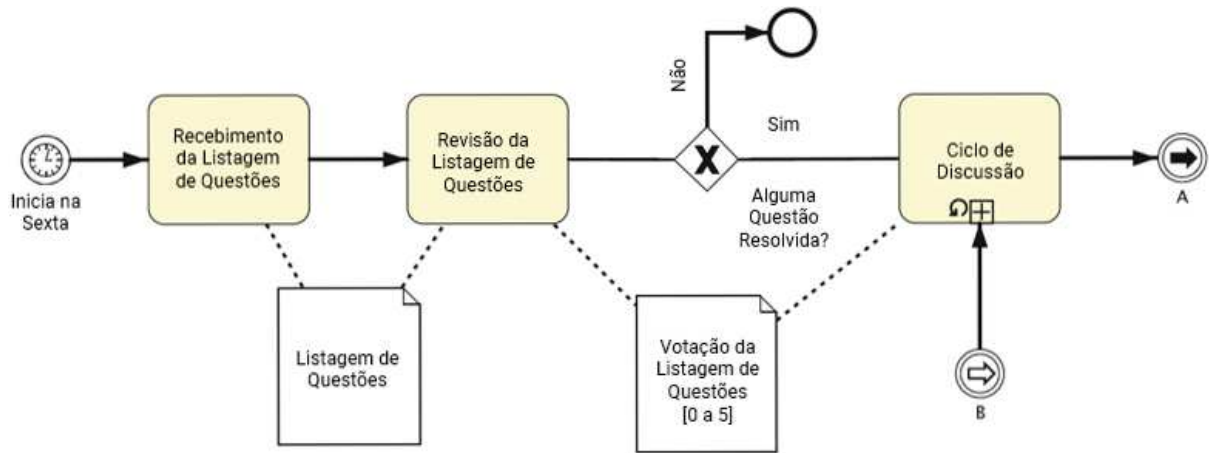
O BPMN (*Business Process Model and Notation* ou Modelo de Processo de Negócios e Notação) é o resultado de um acordo celebrado com o objetivo de padronizar uma notação a ser utilizado pelos fornecedores de soluções que envolvam modelos de processos de negócios. Esta notação gráfica descreve os processos de negócios, de forma intuitiva, sendo capaz de representar estruturas complexas de processos, e o seu conteúdo apresenta mais informações e detalhamento do que as redes de Petri (BURATTIN, 2015).

Os componentes de um diagrama BPMN, de acordo com Burattin (2015) são:

- eventos: define-se como os acontecimentos que ocorrem durante o processo, possuem uma causa e um resultado. Os eventos são representados por círculos, e podem ser do tipo: início, intermediário ou final;
- atividades: identifica o trabalho realizado por uma empresa, são representados por retângulos arredondados, podem ser tarefas ou subprocessos;
- entradas: estrutura utilizada para controlar as divergências e convergências do fluxo do processo;
- sequência e fluxos de mensagens e associações: conectores entre os componentes do gráfico, e indicam a ordem das atividades, os fluxos de mensagens mostram o envio e o recebimento das mensagens entre os participantes, e por fim as associações são utilizadas para conectar artefatos com outros elementos do gráfico.

A Figura 6 apresenta um exemplo de um diagrama BPMN.

FIGURA 6 – EXEMPLO DE BPMN



FONTE: Traduzida de Burattin (2015)

Apresentados os conceitos envolvendo a mineração de processos, em sequência conceitua-se a fraude.

2.2 FRAUDE

Os desvios ou anomalias podem estar presentes em processos de negócios originados por erros ou variações na operação, porém em alguns casos as anomalias podem ter origem em comportamentos fraudulentos, realizados das mais diversas formas e ocasionando perdas significativas (SARNO, SINAGA e SUNGKONO, 2020). As fraudes podem ser executadas por um indivíduo isolado ou por grupos, sendo uma temática de grande interesse tanto na literatura profissional quanto na acadêmica, que devem ser estudadas de formas categorizadas, de modo a diferenciar os variados tipos de crimes, devido as suas características peculiares (MAULIDI, 2020).

O comportamento vinculado a ilegalidade frequentemente é atribuído às questões de traços de personalidade ou da vinculação a fragilidade moral dos indivíduos, além de variáveis não relacionadas ao indivíduo, mas vinculadas ao ambiente contextual e situacional em que se encontram (MAULIDI, 2020). A fraude sempre irá ocorrer onde existir uma oportunidade, e o crescimento da fraude não diminuirá a menos que se adote medidas preventivas sejam adotadas em todos os níveis hierárquicos da organização (BIEGELMAN, BARTOW, 2012).

A prevenção da fraude requer um sistema de regras com o objetivo de minimizar a sua ocorrência, a prevenção deixa de ser apenas uma boa prática comercial, sendo essencial e necessária atualmente, devido aos riscos que as organizações expostas e as consequências podem ocasionar. Os riscos surgem pela responsabilização da organização pelos atos cometidos pela própria organização ou pelos funcionários que agiram em nome dela, além dos riscos financeiros com as perdas pelo ato lesivo, ações judiciais de acionistas, multas e processos judiciais. Nesta lista, agregam-se os riscos da imagem da organização, que pode ser ocasionado pelos atos fraudulentos (BIEGELMAN, BARTOW, 2012).

Outro ponto importante na compreensão da fraude é conhecer as causas que originam esses comportamentos fraudulentos, com isso subsidiar na concepção de intervenções de prevenção e detecção de fraudes (MAULIDI, 2020). Algumas das teorias que apresentam e justificam as motivações desses comportamentos, são apresentadas no tópico seguinte, após a caracterização de *outlier* e anomalias e em que aspectos a gestão da informação se aproxima da fraude.

2.2.1 Motivação da fraude

A motivo que levam as pessoas a cometerem a fraude em muitos casos é de difícil compreensão, Biegelman e Bartow (2012) apresentam as teorias do Triângulo da Fraude, Escala de Fraude e Diamante da Fraude que buscam determinar as causas que originam este tipo de comportamento. Acrescentam-se as abordagens de Desai (2020) que descreve modelos teóricos que expliquem as causas das fraudes corporativas, que incorporam fatores pessoais e financeiros, como o M.I.C.E. e o modelo ABC:

- a) a Teoria do Triângulo da Fraude desenvolvida pelo Dr. Donald Cressey conclui que são necessários três elementos para que o indivíduo fique propenso a cometer uma fraude: o motivo, a oportunidade e a racionalização. Cada um destes elementos é necessário e se inter-relaciona, e a ausência de um não permitiria que a pessoa cometesse a fraude. O motivo costuma estar relacionado com a pressão financeira, ainda este elemento implica em ações ou reações vinculadas a uma emoção ou desejo. A oportunidade é a circunstância ou ambiente favorável

que permite que ocorra a fraude, muitas vezes relacionado ao cargo ou posição na hierarquia da empresa, e o terceiro elemento, a racionalização é como o fraudador justifica a ação inadequada, muitas vezes motivada por uma necessidade financeira, e beneficiado pela oportunidade;

- b) a Escala da Fraude se assemelha ao triângulo da fraude, desenvolvido pelo Dr. W. Steve Albrecht teoriza que mesmo estando presentes a oportunidade e a necessidade, alguns funcionários nunca cometerão fraude, sendo que a motivação para cometer a fraude dependerá de uma complexa combinação dos três fatores: oportunidade, motivação e racionalização. Em seu estudo, identificou que as pessoas que cometiam as fraudes viviam além do que sua renda permitia, com dívidas pessoais ou de jogo, às vezes a motivação vinha do desejo de melhoria do padrão social ou pela pressão da família, fatores como acreditar que eram pagos ou subvalorizados também geravam uma propensão ao crime;
- c) o Diamante da Fraude, é uma evolução do triângulo da fraude, adiciona um quarto elemento: a capacidade. Este elemento inclui a personalidade e as características do fraudador, onde inclui-se o conhecimento, a criatividade, a autoconfiança e o ego, neste contexto, uma pessoa pode ter a oportunidade de cometer a fraude, porém não possui a capacidade de realizar a fraude;
- d) o M.I.C.E., é um acrônimo das palavras em inglês *money*, *ideology*, *coercion* e *ego*, traduzindo, dinheiro, ideologia, coerção/compulsão e ego, e a sigla M.I.C.E. remete a palavra ratos. Este modelo propõe que nem todas as fraudes corporativas são motivadas por questões de pressões financeiras ou oportunidades financeiras, levando para questões de prestígio social ou situações culturais de competição, em especial nos casos dos crimes de colarinho branco. O ambiente de uma estrutura fraca de governança, aliada a ganância por dinheiros, uma concepção ideológica voltada a corrupção, o poder de coerção, a senso do direito a cometimento do ato ilícito e o ego inflado, são o que motivam fraudadores segundo o M.I.C.E., que se caracterizam por ação de um predador, ao

contrário do descrito no triângulo da fraude que indicam ações de um fraudador acidental;

- e) o modelo A.B.C. proposto por Ramamoorti et al. (2009 apud DESAI, 2020) caracteriza a fraude segundo a frase, do inglês, “*a bad Apple, a bad Bushel, and a bad Crop*”, uma tradução literal, “uma maçã podre, um cesto podre, uma plantação podre”. Neste contexto, a maçã podre refere-se ao fraudador, o cesto refere-se as pessoas que convivem com o fraudador, muitas vezes coniventes com as praticadas, e a plantação podre, remete as questões culturais e sociais, que criam um ambiente com baixa ênfase a ética e legalidade.

Um dos fatores de crimes financeiros é a dificuldade dos proprietários e partes interessadas monitorarem a gestão das empresas, confiada em muitas vezes à terceiros, permitindo o oportunismo de ações fraudulentas. Esta relação a ocorrência da fraude corporativa surge quando existe uma “violação de tal confiança onde a administração não atua no melhor interesse da corporação partes interessadas, mas atua de forma oportunista para beneficiar certas partes interessadas titulares às custas de outros” (DESAI, 2020, tradução minha).

Conceituado algumas teorias que motivam a atividade da fraude, na sequência apresentam-se aspectos relacionais envolvendo elementos que possibilitam a identificação de *outliers*, anomalias e fraudes.

2.2.2 *Outliers* ou anomalias em processos

Algumas características definem um objeto ou processo como sendo um *outlier*, quando ocorre um desvio do comportamento normal esperado, a partir de um conjunto de dados conhecidos, quando se apresenta valores longes dos valores médios ou esperados, isto é, quando apresentam alguma anormalidade, desta forma são objetos interessantes para a análise. O *outlier* pode ser caracterizado quanto ao tamanho, abrangendo um único objeto, subconjuntos, grupos ou uma rede inteira, quanto a sua diversidade, pela frequência de ocorrência ou a forma como afeta o desempenho de algoritmos ou sistemas (RANGA SURI; NARASIMHA MURTY; ATHITHAM, 2019).

Identificar um verdadeiro *outlier* de ruídos em conjuntos de dados é uma tarefa muitas vezes difícil, pelo viés que ambos indicam um desvio do que seria uma normalidade para os dados.

[...] a questão do que constitui um desvio suficiente para um objeto ser considerado como um outlier é muitas vezes um julgamento subjetivo com base no que desperta a atenção do analista. Embora o ruído em si possa não ser interessante para o analista, sua identificação e remoção continua a ser uma tarefa importante na mineração de dados. Assim, tanto o ruído quanto os problemas de detecção de valores discrepantes são igualmente importantes para uma análise de dados eficaz. Nesse contexto, a melhor maneira de descobrir outliers e distingui-los do ruído é usar o *feedback* de exemplos atípicos de interesse anteriormente conhecidos. (RANGA SURI; NARASIMHA MURTY; ATHITHAM, 2019, p.8, tradução minha)

A identificação de processos raros, compreende tarefas como a descoberta de eventos desconhecidos e a detecção de instâncias dentro de um conjunto de dados, os eventos de interesse caracterizam-se não por serem raros, mas por não terem uma explicação de acordo com os comportamentos reais. A caracterização do evento como sendo simplesmente uma coincidência ou um ato malicioso, depende da sua inserção em contexto e da explicação do seu acontecimento, desta forma um evento de interesse relaciona-se a definição de anomalia, podendo ou não ser um *outlier* (RANGA SURI; NARASIMHA MURTY; ATHITHAM, 2019).

A detecção de *outliers* busca a identificação de um objeto ou conjuntos de dados minoritários que se desvirtuam significativamente da população analisada, sua precisão e eficiência é sempre um desafio, considerando que a incumbência ao analista de definição do que seria um *outlier*, e relacionando ao quão diferente, excepcional ou inconsistente são estes dados. Algumas aplicações de mineração de dados procuram a minimização ou eliminação destes *outliers*, confundindo com ruídos no conjunto de dados, e muitas vezes podendo resultar em perdas de importantes informações, pois estes *outliers* podem ser de interessantes, como no exemplo de detecção de fraudes, onde podem indicar atividades fraudulentas (RANGA SURI; NARASIMHA MURTY; ATHITHAM, 2019).

Em muitas pesquisas são identificados atributos e padrões que podem descrever as fraudes que ocorrem nos processos de negócios. Sarno, Sinaga e Sungkono (2020) relatam seis tipos de atributos que identificam situações anômalas ou comportamentos fraudulentos em processos de negócios:

- a) atividade ignorada (*skipped activity*): como o próprio termo descreve, ocorre a anomalia quando atividade é ignorada dentro de um procedimento operacional padrão, ou seja, numa sequência de atividades pula-se a atividade constante no procedimento padrão. Este atributo pode ser do tipo: sequência ignorada, quando se ignora uma atividade constante de uma sequência, ou do tipo decisão ignorada, quando se ignora uma atividade que necessita de uma decisão ou autorização para prosseguir para a outra etapa;
- b) tempo processamento incorreto (*wrong throughput time*): a anomalia ocorre quando a atividade é realizada num tempo menor ou maior do que os limites estabelecidos no procedimento operacional padrão. É dividido em dois tipos: abaixo do tempo mínimo, ou quando o tempo de processamento máximo está incorreto;
- c) recurso incorreto (*wrong resource*): o recurso descrito refere-se a pessoa responsável pela execução da atividade, a ocorrência da tipologia de recurso incorreto refere-se à situação em que uma atividade não é executada pelo funcionário autorizado para realizar a atividade de acordo com as funções estabelecidas no procedimento operacional padrão;
- d) função errada (*wrong duty*): É uma condição que ocorrer quando um funcionário executa duas ou mais atividades diferentes em um mesmo processo em execução, podendo ser de três tipos: função sequencial errada, o funcionário executa mais de uma atividade na sequência; função errada de decisão, o funcionário toma decisão sem ter a autorização para tomá-la; e a função errada combinada, é um conjunto das outras duas;
- e) padrão incorreto (*wrong pattern*): Ocorre esta situação quando uma sequência de atividades é executada diferente da sequência de atividades constantes no procedimento operacional padrão;

- f) decisão errada (*wrong decision*): Considerando um processo de tomada de decisão pré-estabelecido, esta tipologia ocorre quando uma decisão é feita em desacordo com o padrão, não está em conformidade.

Como forma de identificar as tipologias de anomalias Sarno, Sinaga e Sungkono (2020) exemplificam quatro tipos de análises de mineração de processos:

- a) análise de fluxo de controle. Esta análise pode ser executada manualmente ou de forma automática, permitindo a detecção de fraudes nos casos de atividade ignorada ou executada num padrão incorreto. Podem ser aplicados métodos de verificação de conformidade, com o objetivo de comparar processos em execução com processos operacionais padronizados. As anomalias podem ser identificadas em uma ou mais eventos dentro do processo, os resultados encontrados também podem indicar comportamentos fraudulentos;
- b) análise do recurso responsável. Cada evento possui uma função ou atribuição específica dentro de um processo, comparando estas atribuições de um processo em execução com um procedimento operacional padrão, pode-se obter a probabilidade de ocorrência de fraude em relação aos recursos empregados;
- c) análise de tempo de processamento. A análise do tempo de transferência ou processamento pode ser executada medindo-se o intervalo de tempo entre as atividades, que pode ser comparado com o procedimento operacional padrão, distorções nos tempos de execução das atividades em relação ao padronizado indicam ocorrências de anomalias;
- d) análise do ponto de decisão. Nesta análise procura-se identificar a existência de anomalias no ponto de decisão, ou seja, descobrir se o responsável pela decisão é o funcionário autorizado para exercer esta função.

Caracterizado os elementos e as situações relacionados a fraude, como a gestão da informação aproxima-se deste relacionamento como fator presente, de forma a agregar soluções aos mecanismos existentes.

2.2.3 Gestão da informação e a fraude

As organizações, sejam de pequeno ou grande porte, utilizam-se de modelos de processos como forma de gerenciar, coordenar, monitorar e validar as atividades, que atualmente são apoiadas por algum sistema de tecnologia da informação. Estes sistemas possibilitam os registros das ações constantes na execução dos processos, que antes realizados manualmente (MUNOZ-GAMA, 2016). A complexidade e a importância da tecnologia da informação nas empresas deixam de ser apenas uma questão técnica, aparecendo uma necessidade da governança, incluindo ações de gestão de riscos e conformidade (*compliance*), que apresenta como principal desafio a integração destas três áreas objetivando a eficácia e eficiência (DE SMET; MAYER, 2016).

Alguns requisitos voltados a gestão de riscos estão incorporados nas organizações, devido normas e regulamentos relacionados à segurança, da mesma forma ocorre na questão da governança e conformidade, diretamente ligados a responsabilidade e funções dos gestores (DE SMET; MAYER, 2016). As recomendações do estudo de Smet e Mayer (2016), no aspecto de governança e gerenciamento de riscos em tecnologia da informação, partem da melhoria da integração destes três elementos, cuidando de especificidades e críticas na segurança da informação, além da padronização e utilização de modelos referenciais.

As técnicas de verificação conformidade forense, a partir de dados históricos, após o término do evento, a necessidade de análises num intervalo menor de tempo e durante a própria execução do processo, produz novos desafios computacionais, concentrando as análises na busca de desvios e suas causas. As técnicas de verificação em tempo real destinam-se especificamente as ações de monitoramento, proteção contra falhas em sistemas, e conseqüentemente na governança, gestão de riscos e verificação de conformidade (MUNOZ-GAMA, 2016).

Algumas das aplicações envolvendo a identificação de *outliers* são listados por Ranga Suri, Narasimha Murty e Athitham (2019).

- identificação de notícias falsas (*fake news*) e comportamento estranhos em sites e redes sociais;
- detecção de anomalias;
- localização de irregularidades no consumo elétrico em redes inteligentes;
- análise e diagnóstico imagens médicas, identificando por exemplo tumores;
- previsão de irregularidades em negociações no mercado de ações;
- avaliação do potencial de ameaças internas nas organizações;
- identificação de falsos vendedores em comércio eletrônico.

Os fatores, situações e ambientes onde reside as atividades fraudulentas são diversos, conceituando-se estes elementos que compõem as peças de um quebra-cabeça, prossegue-se para a definição do roteiro metodológico a ser aplicado.

3 ENCAMINHAMENTOS METODOLÓGICOS

A pesquisa tem abordagem qualitativa. Neste tipo de abordagem considera e existência de “uma relação dinâmica entre o mundo real e o sujeito, isto é, um vínculo indissociável entre o mundo objetivo e a subjetividade do sujeito que não pode ser traduzido em números” (PRODANOV; FREITAS, 2013, p.70). Os dados são coletados através de interações sociais e analisados subjetivamente pelo pesquisador, isto é, o “fenômeno remete-nos à interpretação de um fato feita por um observador” (APPOLINÁRIO, 2011, p.150).

Na temática de pesquisa, a abordagem qualitativa estará presente no desenvolvimento dos três objetivos específicos, primeiramente na revisão da literatura através da identificação de características constantes em estudos relacionados, na sequência no mapeamento comparativo das ferramentas de mineração de processos, também identificando características qualitativas, e por fim uma análise do pesquisador dos resultados da aplicação em situação simulada, o que caracteriza a pesquisa como aplicada.

Quanto à natureza a pesquisa é aplicada, onde o objetivo é a resolução de um problema concreto e imediato da humanidade (APPOLINÁRIO, 2011). Do ponto de vista dos objetivos a pesquisa é descritiva, quando o pesquisador descreve e interpreta a realidade, sem nela interferir, não estabelece relações de causalidade (APPOLINÁRIO, 2011; PRODANOV; FREITAS, 2013).

Do ponto de vista dos procedimentos é bibliográfica, considerando a análise sendo elaborada a partir outras pesquisas, com o intuito de avaliar determinado campo de pesquisa (PRODANOV, FREITAS, 2013, p.54). Além disso, é uma pesquisa do tipo *ex post facto*, que se caracteriza que emprega uma seleção de dados coletados por estudos anteriores, o experimento é realizado depois de ocorridos os fatos (APPOLINÁRIO, 2011, p.148).

A pesquisa caracteriza-se pela fundamentação em estudos e dados históricos com o intuito de fundamentar a aplicação da mineração de processos, as possibilidades em termos de opções ferramentais e a análise de dados abertos, dados que espelham processos reais na gestão pública. A Figura 7 resume a caracterização da pesquisa, quanto à abordagem, natureza, objetivos e meios.

FIGURA 7 – CARACTERIZAÇÃO DA PESQUISA



FONTE: Elaborada pelo autor (2020).

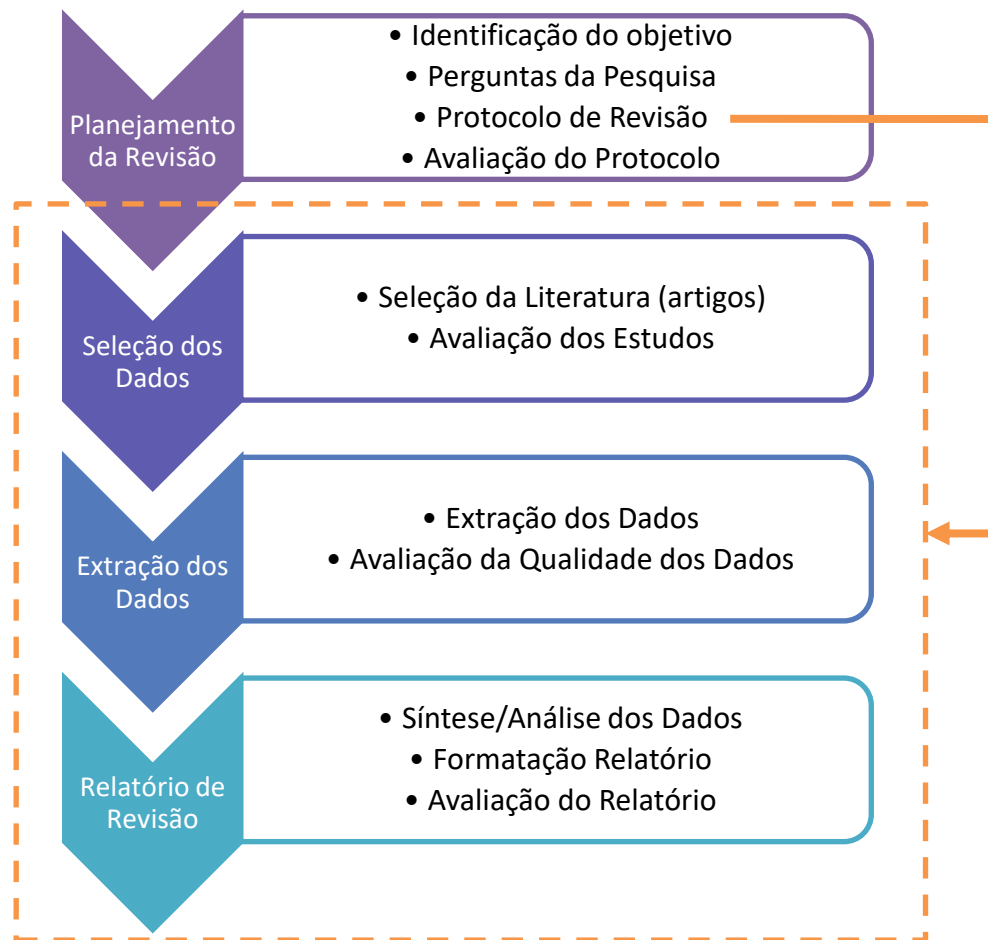
Na sequência as subseções apresentam a metodologia utilizada para cada objetivo específico, iniciando-se pela revisão da literatura.

3.1 REVISÃO SISTEMÁTICA DA LITERATURA

A coleta e análise de dados orientam-se nos trabalhos de Kitchenham e Charters (2007) e de Okoli (2019). Ambos descrevem as diretrizes de planejamento e execução de uma revisão sistemática da literatura, que será empregada neste estudo.

A revisão sistemática da literatura tem o objetivo de condensar as evidências existentes sobre determinado tema, resumindo benefícios e limitações, identificando lacunas em pesquisas ou sugerindo novas áreas de investigação (KITCHENHAM, CHARTERS, 2007). A Figura 8 apresenta o roteiro da revisão da literatura, em linhas gerais:

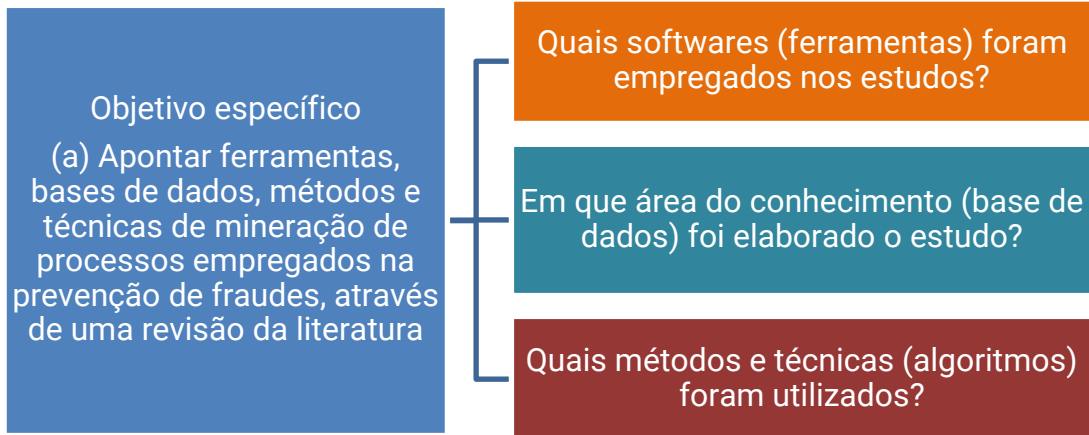
FIGURA 8 – ROTEIRO DA REVISÃO DA LITERATURA



FONTE: Adaptada de Kitchenham, Charters (2007) e Okoli (2019).

A identificação do objetivo revela o propósito da revisão de literatura de atender aos requisitos de rigor acadêmico e ao mesmo tempo em que produz valor (OKOLI, 2019). Quanto a pergunta de pesquisa tem seu relacionamento aos propósitos do primeiro objetivo específico, que busca clarear os direcionamentos dos estudos publicados e embasar o presente estudo. A Figura 9 descreve o objetivo desmembrado em três questões focado na solução da ferramenta de mineração de processos, na área em que a ferramenta foi aplicada, e quais os meios, em termos de algoritmos ou modelos, utilizados no desenvolvimento de exemplos, estudos de casos e testes.

FIGURA 9 – QUESTÕES A SEREM RESPONDIDAS NA REVISÃO DA LITERATURA



FONTE: Elaborada pelo autor (2021).

O Protocolo de Revisão tem o intuito de reduzir a arbitrariedade do pesquisador, a fim de garantir consistência na execução da revisão (KITCHENHAM; CHARTERS, 2007; OKOLI, 2019). O Quadro 2 apresenta o protocolo de revisão adotado nesta pesquisa.

QUADRO 2 – PROTOCOLO DE REVISÃO

Etapas	Passo a Passo	Descrição	Método/Resultado
Perguntas de pesquisa	Abordagem do problema	Apontar ferramentas, bases de dados, métodos e técnicas de mineração de processos empregados na prevenção de fraudes (Figura 9).	
Procedimento de seleção (Critérios)	Estratégia de pesquisa Busca dos artigos Tratamento dos resultados	Definição da estratégia de pesquisa.	em inglês “process mining” e “fraud” em português “mineração de processos” e “fraude”
		Base de dados	3 bases: Scopus, Web of Science e EbscoHost
		Agrupamento dos artigos,	70 artigos encontrados
		remoção de duplicatas ou exclusão, identificação dos artigos completos elegíveis para análise, e inclusão de outros artigos encontrados.	34 artigos duplicados e 1 sem indicação de autores
Extração de dados	Codificação	Leitura dos artigos, identificação das respostas das questões levantadas e mapeamento das repostas.	35 artigos selecionados
Síntese dos dados			
Análise dos dados	Análise		
Relatório	Elaboração do relatório	Relatório em forma de quadro (mapa)	35 artigos selecionados

FONTE: Elaborado pelo autor (2021).

Os resultados encontrados no cumprimento deste objetivo específico refletem nos encaminhamentos do segundo objetivo específico de comparar as ferramentas de mineração de processos disponibilizados no mercado que possibilitem a sua implementação na mitigação de fraudes.

3.2 PROTOCOLO PARA COMPARAÇÃO DAS FERRAMENTAS

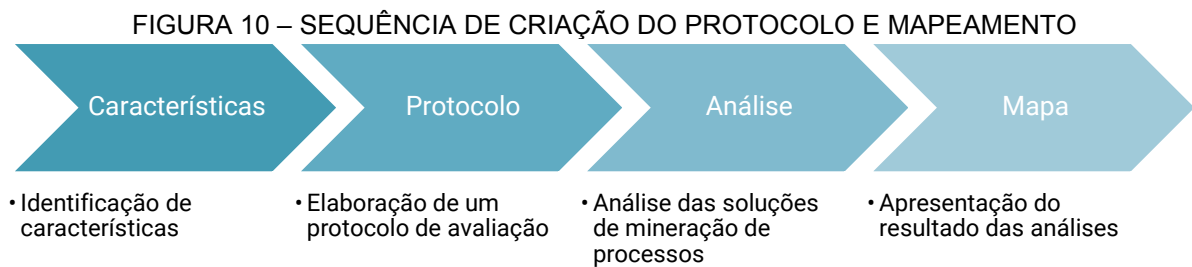
Inicia-se esta subseção relatando alguns dos trabalhos relacionados a comparação de ferramentas de mineração de processos (Quadro 3).

QUADRO 3 – TRABALHOS RELACIONADOS A PESQUISA

Autores	Ferramentas de Mineração de Processos Comparadas
Ailenei (2011)	ARIS Process Performance Manager, Fourspark - Flow, Futura Process Intelligence e QPRr
Verstraete (2014)	ProM, Disco, Celonis, Perceptive PM, QPR, ARIS, Fujitsu, XMANalyzer e StereoLOGIC Discovery Analyst
Agarwal, Singh (2014)	ProM, Disco, EmiT Tool e Aris PPM
Kebede (2015)	ProM 6.4.1, Disco 1.8.2 e Celonis
Ramos (2019)	ProM 6.8, Apromore 7.0.1, BupaR 0.4.0.900, Disco, Celonis ARIS PPM, QPR - Process Analyzer
Bru, Claes (2018)	Disco, ProM, Celonis, Minit, Aris, myInvenio, QPR, Process Gold, Fujitsu, Appian
Çelik, Akçetin (2018)	ProM 6.5.1, Disco 1.9.5, Celonis e myInvenio
Devi, Kumudavalli, Sudhamani (2017)	ProM 6.6, Disco Enterprise Edition e Celonis Enterprise Edition
Viner, Stierle, Matzner (2020)	ABBYY, Apromore, Aris, Optix, Celonis, Disco, EverFlow, Lana PM, Logpickr, Mehrwerk PM, MiniT, myInvenio, PAFnow, ProDiscovery, QPR, Signavio e UIPath PM
Drakoulogkonas, Apostolou (2021)	Apromore Community Edition, Celonis, Disco, myInvenio, ProM

FONTE: Elaborado pelo autor (2021).

Para o desenvolvimento do estudo considerou-se a sequência dos trabalhos para comparação das ferramentas de mineração de processos, que começa com a identificação das características, elaboração de um protocolo com as características adotadas para avaliação, a análise dos achados e concluindo com a planificação dos resultados, conforme Figura 10.



FONTE: Elaborado pelo autor (2021).

As características incluídas no protocolo de comparação tiveram com referências os estudos de Van der Aalst (2016), Ramos (2019) e Viner, Stierle e Matzner (2020). Agrupado em três blocos de características: especificações gerais das ferramentas, os tipos de arquivos suportados para importação e exportação dos arquivos contendo os registros de eventos e as características específicas relacionadas aos métodos e técnicas envolvendo a mineração de processos (Figura 11).

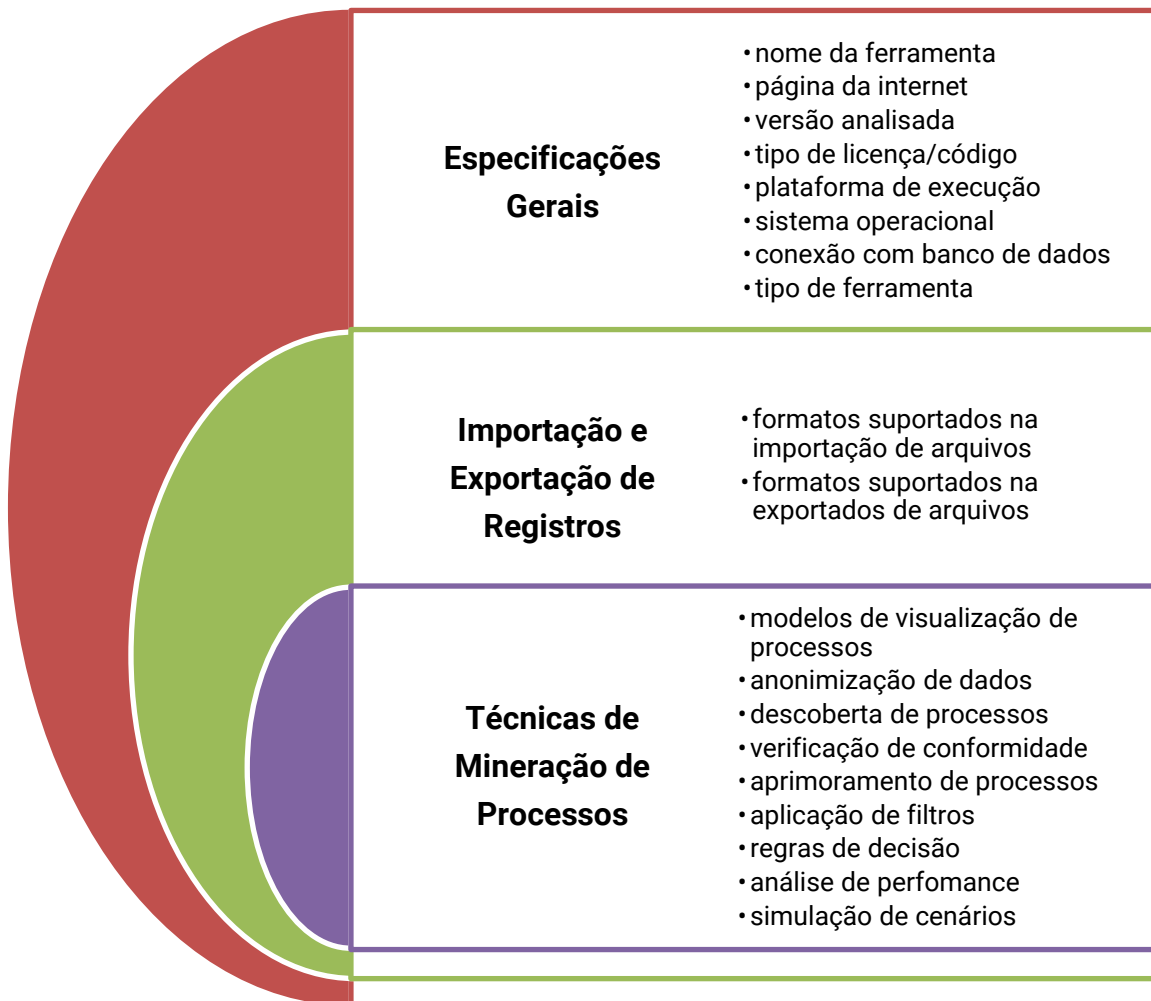
Nas especificações gerais consta o detalhamento da versão da ferramenta de mineração de processos, devido as diferenças existentes entre as versões ou atualizações, que podem incluir ou excluir funcionalidades. Quanto à categoria da ferramenta de mineração de processos Van der Aalst (2016) descrever dois tipos:

- software dedicado – é uma solução específica para a aplicação de mineração de processos, para a análise de dados e processos;
- software embutido – a aplicação da mineração de processos está incluída como funcionalidade de outros tipos de soluções: BPMS², ERP³, *business intelligence* ou soluções de mineração de dados.

² BPMS - Business Process Management System ou Sistema de Gestão de Processos de Negócio.

³ ERP - Enterprise Resource Planning ou Planejamento de Recursos Empresariais.

FIGURA 11 – CARACTERÍSTICAS IDENTIFICADOS NAS SOLUÇÕES DE MINERAÇÃO DE PROCESSOS



FONTE: Elaborada pelo autor (2021).

Ainda, segundo Van der Aalst (2016) as ferramentas podem ser de código aberto ou de código fechado:

- software de código aberto – quando a solução de mineração de processos pode ser alterada ou estendida, com disponibilidade pública;
- software de código fechado – em geral as soluções comerciais ou proprietárias que não os códigos fonte não são públicos, não permitindo alteração ou extensão.

Complementando o primeiro bloco de características, além da categoria da ferramenta de mineração de processos e do tipo de código, também serão identificados (RAMOS, 2019; VINER, STIERLE e MATZNER, 2020):

- a página onde está disponível a ferramenta, para compra ou *download*, caso possua versão gratuita, acadêmica ou de teste;
- a versão analisada, considerando que alterações na ferramenta com a inclusão ou exclusão de recursos;
- plataforma de execução, como plataformas na nuvem (cloud), software como serviço (*SaaS – Software as a Service*) ou aplicativo instalado no computador do usuário (*stand-alone*);
- sistema operacional, referente a versão para download e instalação no computador;
- conexão com banco de dados, conectores com drivers do tipo ODBC (*Open Database Connectivity*) ou JDBC (Java Database Connectivity);

O segundo bloco de características compreende a importação e a exportação de dados, no aspecto do tipo de arquivo, como por exemplo: CSV, XLSX, XES, MXML, além de extensões de imagem como JPEG ou PNG. O terceiro bloco de características depreende os recursos específicos da técnica de mineração de processos (RAMOS, 2019; VINER, STIERLE e MATZNER,2020):

- modelo de visualização do processo;
- anonimização de dados;
- descoberta de processos;
- verificação de conformidade;
- aprimoramento do processo;
- aplicação de filtros;
- análise de redes organizacionais (sociais);
- regras de decisão;
- análise de performance;
- simulação de cenários;

O protocolo terá como referência as características apontadas na Figura 12, com as análises fundamentais nas informações disponibilizadas nas páginas das ferramentas, de trabalhos relacionados, e da utilização das soluções disponibilizados gratuitamente para utilização ou em versão temporária acadêmica.

O mapa comparativo foi executado a partir da leitura dos artigos, primeiramente para a formação das principais características das ferramentas e posteriormente na identificação em qual grupo as ferramentas se enquadrava. A escolha das ferramentas analisadas levou em consideração a recorrência nos estudos, restringindo-se a análise de 8 ferramentas de mineração de processos.

Considerando as diferenciações entre versões da mesma ferramenta, buscou-se identificar a versão analisada tanto nas características apresentadas nos estudos, como diretamente nas páginas onde as ferramentas estavam disponibilizadas, além de alguns testes nas próprias ferramentas utilizadas. Os resultados foram transcritos em forma de três quadros para que o leitor pudesse facilmente realizar uma análise comparativa, não somente para os objetivos deste estudo, mas para atender outras necessidades.

3.3 SIMULAÇÃO DO EMPREGO DA MINERAÇÃO DE PROCESSOS

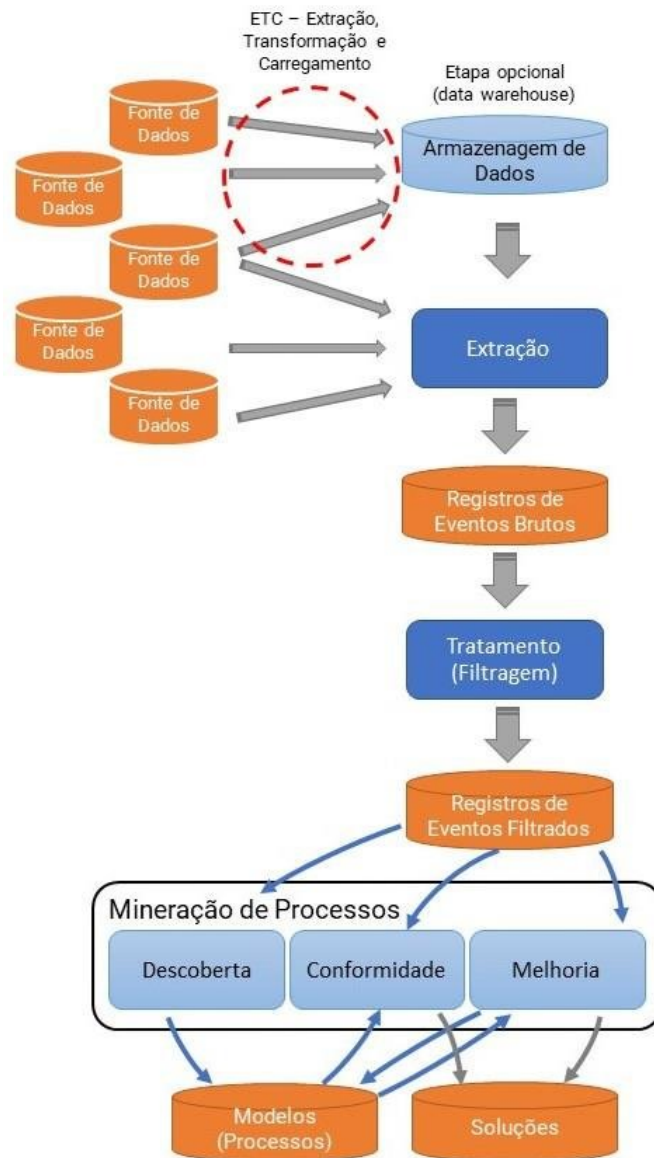
O desenvolvimento do último objetivo específico de “selecionar e julgar o emprego da mineração de processos em um conjunto de registros de licitações de obras públicas de municípios paranaenses” será realizado em duas etapas. A primeira etapa será a elaboração de um fluxo de trabalho para o tratamento da base de dados que subsidiará as simulações a serem realizadas em ferramentas de mineração de processos, e a segunda etapa compreende a utilização das ferramentas, verificação e análise dos resultados pelo pesquisador.

3.3.1 Tratamento da base de dados

A Figura 12 apresenta o fluxo de trabalho para obtenção dos dados de diferentes fontes, etapa preliminar que deve ser efetuada antes da aplicação da ferramenta de mineração de processos. Um conjunto de registros de eventos (*event logs*) adequados é essencial para a mineração de processos. Surge o desafio de extrair dados de diferentes fontes e a análise da qualidade destes dados, que

posteriormente são filtrados de acordo com os objetivos propostos (VAN DER AALST, 2016).

FIGURA 12 – EXTRAÇÃO E TRATAMENTO DOS DADOS PARA ANÁLISE



FONTE: Adaptada de Van Der Aalst (2016).

As fontes de dados podem ser estruturadas ou não estruturadas, compostas por conjuntos bem definidos de metadados ou terem metadados importantes faltantes, e originadas de páginas da *web*, *e-mails*, textos digitalizados, entre outras. Mesmo em dados estruturados e metadados bem descritos, a complexidade do sistema pode ser um fator que pode dificultar a extração dos dados. A opção é transformar os dados conforme a necessidade (VAN DER AALST, 2016). A Figura 13 descreve as etapas

da construção da base de dados composta pelos registros de licitações de obras públicas de municípios paranaenses, e disponibilizados em plataforma de dados abertos do Tribunal de Contas do Estado do Paraná.



FONTE: Elaborada pelo autor (2021).

Na construção da base de dados utilizaram-se os campos:

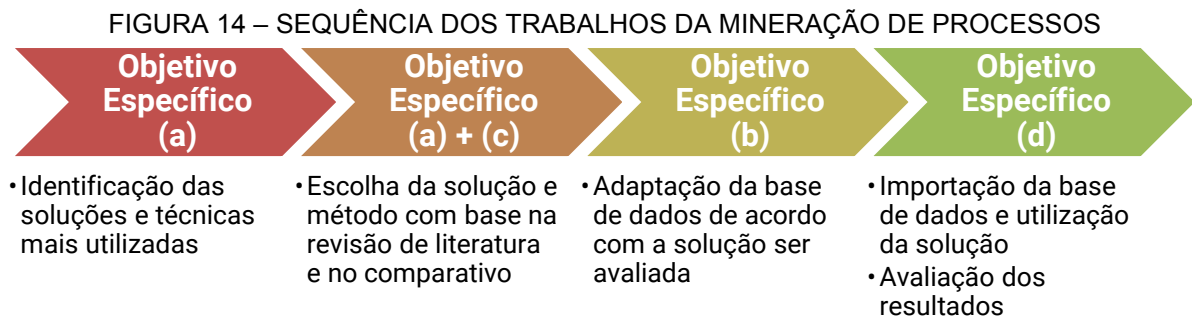
- IdIntervencao;
- NrAcompanhamento;
- DtAcompanhamento;
- DsTipoAcompanhamento; e
- dsTipoCategoriaBem.

Estes campos estão contidos nos arquivos atualizados até 15/02/2021 (ANEXO I):

- “acompanhamentos_base_de_dados.csv” e;
- “bens_patrimoniais_base_de_dados.csv”.

3.3.2 Análise dos resultados da utilização de ferramentas de mineração de processos

O quarto objetivo condensa os três primeiros objetivos específicos e tem o intuito de “aplicar o método de mineração de processos na base de dados construída”. A Figura 14 apresenta a sequência dos trabalhos até a obtenção do artefato para ser avaliado.



FONTE: Elaborada pelo autor (2021).

Para avaliação dos resultados serão utilizados os relatam seis tipos de atributos que identificam comportamentos fraudulentos em processos descritos no trabalho de Sarno, Sinaga e Sungkono (2020). Os atributos são:

- atividade ignorada (*skipped activity*);
- tempo processamento incorreto (*wrong throughput time*);
- recurso incorreto (*wrong resource*);
- função errada (*wrong duty*);
- padrão incorreto (*wrong pattern*).

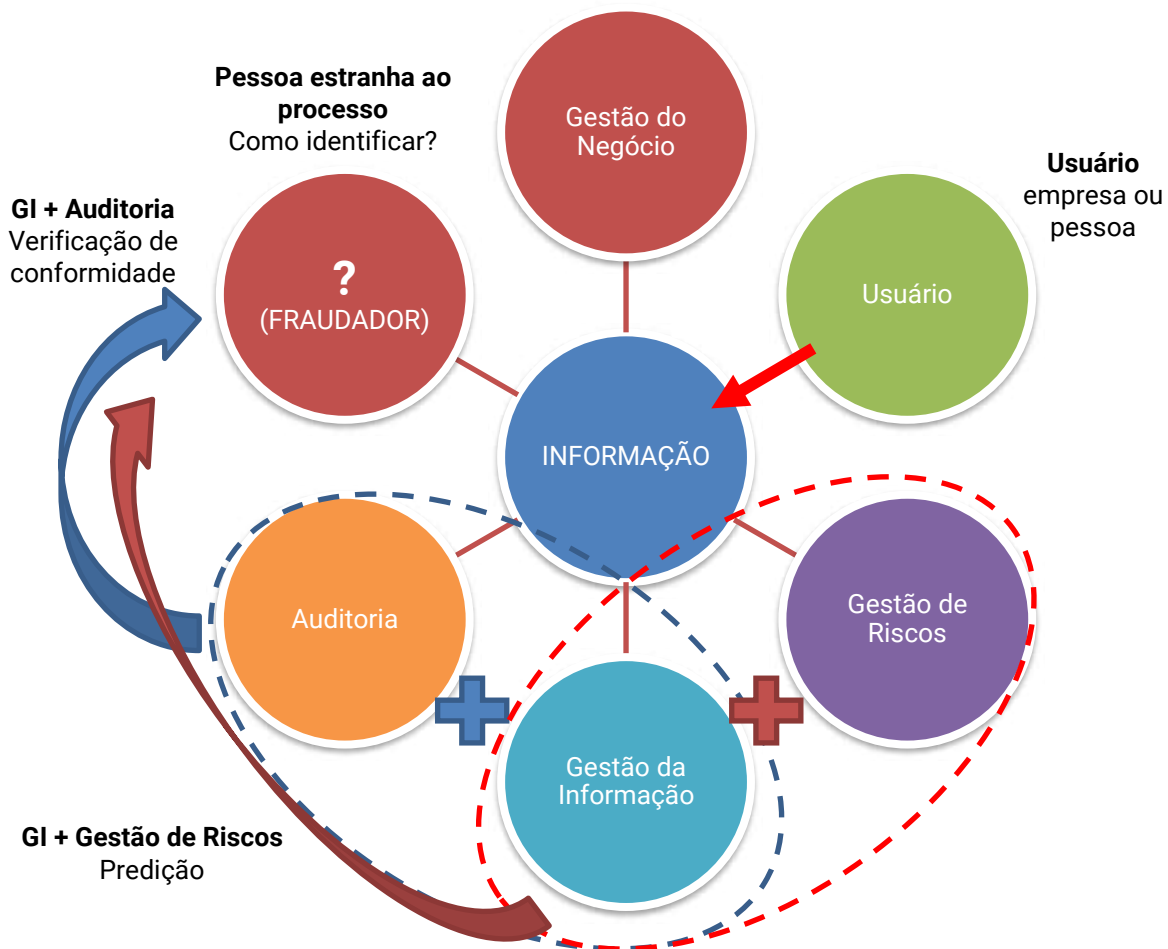
Na próxima seção são apresentadas as análises e discussões dos resultados, considerando os encaminhamentos metodológicos descritos.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesta seção são apresentados os resultados deste estudo, bem como se discute os mesmos. Inicia-se com as algumas características identificadas no levantamento da literatura em termos da relação entre a mineração de processos e sua aplicabilidade na identificação de fraudes, na sequência apresenta-se um comparativo entre alguns softwares de mineração de processos. Por fim, demonstra-se com base em dados abertos, que contemplam execução de obras públicas de municípios paranaenses, a aplicação de ferramentas de mineração de processos.

A identificação da fraude pode ocorrer de preventiva ou após a ocorrência, a informação e sua gestão localizam-se em todas as etapas, conforme visualizamos na Figura 15.

FIGURA 15 – GESTÃO DA INFORMAÇÃO, AUDITORIA E GESTÃO DE RISCOS NA IDENTIFICAÇÃO DE FRAUDADORES



FONTE: Elaborada pelo autor (2021).

4.1 APONTAMENTOS DA REVISÃO DA LITERATURA

A revisão baseou-se em pesquisa realizada nas bases de dados *Scopus*, *Web of Science* e *EbscoHost* resultando em 35 artigos publicados em periódicos e anais de congressos. Este levantamento foi apresentado na 4ª Conferência sobre Sistemas de Informação na América Latina – ISLA 2020, em 11 de agosto de 2020, e teve como objetivo uma análise bibliométrica da produção científica relacionada a mineração de processos e fraude. O Quadro 4 estão relacionadas as ferramentas mais utilizadas nos estudos que compõem a revisão sistemática da literatura.

QUADRO 4 – FERRAMENTAS MAIS UTILIZADAS NOS ESTUDOS DA REVISÃO SISTEMÁTICA DA LITERATURA

Ferramenta	Número de Artigos	Autor(es)
ProM	14	JALALI, H.; BARAANI, A. (2010), JANS, M. <i>et al.</i> (2011), HUANG, Z.; CONG, Q.; HU, J. (2012), MARDANI, S.; SHAHRIARI, H. R. (2013), BOSE, R. P. J. C.; AALST, W. M. P. (2013), BEZERRA, F.; WAINER, J. (2013), BOSE, R. P. J. C.; VAN DER AALST, W. M. P. (2013), CHAO, O.; CHENG, H.; JUAN, Y. (2014), SARNO, R. <i>et al.</i> (2015), PANE, E. S.; VIBAWA, A. D.; PURNOMO, M. H. (2017), FOLINO, F.; GUARASCIO, M.; PONTIERI, L. (2017), RAHMAWATI, D.; SARNO, R. (2017), CAESARITA, Y.; SARNO, R.; SUNGKONO, K. R. (2017), SARNO, R.; SINAGA, SUNGKONO, K. R. (2020)
ProM e CoBeFra	1	ADRIANSYAH, A <i>et al.</i> (2014)
Disco, ProM e R (pacotes edeaR e bupaR)	1	CHIU, T.; JANS, M. (2017)
Disco	1	ZERBINO, P. <i>et al.</i> (2018)
Celonis Process Mining	1	BAADER, G.; KRCCMAR, H. (2018)
EMiT (Enhanced Mining Tool)	1	VAN DER AALST, W. M. P.; DE MEDEIROS, A. K. A. (2005)
YAWL editor (fluxo de trabalho)	1	RAHMAWATI, D.; YAQIN, M. A.; SARNO, R. (2016)
APD tools	1	GENGA, L. <i>et al.</i> (2017)
Sem ferramenta específica	12	YANG, W.-S.; HWANG, S.-Y. (2006), BEZERRA, F.; WAINER, J. (2011), PAULRAJ, P.; NEELAMEGAM, A. (2013), HUDA, S. <i>et al.</i> (2014), VAN DER AALST, W. M. P. (2014), CUZZOCREA, A. <i>et al.</i> (2017), RAHMAWATI, D. <i>et al.</i> (2017), CUZZOCREA, A. <i>et al.</i> (2018), TAVARES, G. M. <i>et al.</i> (2018), SILVA, F. A. R. (2018), NOLLE, T. <i>et al.</i> (2018), FAZZINGA, B. <i>et al.</i> (2020)

Fonte: Elaborado pelo autor (2021).

Da leitura dos 35 artigos identifica-se a utilização da ferramenta de mineração de processos ProM em 16 dos estudos, sendo 14 utilizando somente esta solução e 2 analisando com outras soluções como o Disco (CHIU; JANS, 2017). O Disco também foi empregado no estudo de Zerbino, et al. (2018), e ainda em 12 estudos não foram empregados nenhuma ferramenta específica.

No Quadro 5 estão listadas as áreas de aplicação em que foram empregadas as técnicas de mineração de processos. As três bases de dados com 5 artigos cada são na área de saúde, onde encontramos, por exemplo, pesquisas referentes a tratamento de câncer (CUZZOCREA, 2017; CUZZOCREA, et al., 2018), na área de processos de aquisição de bens, e aplicações no setor bancário, principalmente envolvendo a solicitação de crédito pessoal (BOSE; AALST, 2013; HUDA, et al., 2014, SARNO, et al., 2015; PANE; VIBAWA; PURNOMO, 2017; RAHMAWATI, et al., 2017; RAHMAWATI; SARNO, 2017).

QUADRO 5 – BASE DE DADOS UTILIZADAS NA MINERAÇÃO DE PROCESSOS

Área de Aplicação	Número de Artigos	Autor(es)
Saúde (hospitais, clínicas)	5	YANG, W.-S.; HWANG, S.-Y. (2006), ADRIANSYAH, A et al. (2014), CHAO, O.; CHENG, H.; JUAN, Y. (2014), CUZZOCREA, A. et al. (2017), CUZZOCREA, A. et al. (2018)
Processos de aquisição de bens (produtos e serviços)	5	JANS, M. et al. (2011), RAHMAWATI, D.; YAQIN, M. A.; SARNO, R. (2016), CAESARITA, Y.; SARNO, R.; SUNGKONO, K. R. (2017), BAADER, G.; KRCMAR, H. (2018), CHIU, T.; JANS, M. (2017),
Bancário (empréstimos, financiamentos, cartão de crédito)	5	BOSE, R. P. J. C.; AALST, W. M. P. (2013), HUDA, S. et al. (2014), SARNO, R. et al. (2015), PANE, E. S.; VIBAWA, A. D.; PURNOMO, M. H. (2017), RAHMAWATI, D. et al. (2017), RAHMAWATI, D.; SARNO, R. (2017)
Eventos aleatórios	4	BEZERRA, F.; WAINER, J. (2011), MARDANI, S.; SHAHRIARI, H. R. (2013), BEZERRA, F.; WAINER, J. (2013), NOLLE, T. et al. (2018)
Gerenciamento operacional	3	FOLINO, F.; GUARASCIO, M.; PONTIERI, L. (2017), TAVARES, G. M. et al. (2018), ZERBINO, P. et al. (2018),
Comércio eletrônico	3	VAN DER AALST, W. M. P.; DE MEDEIROS, A. K. A. (2005), PAULRAJ, P.; NEELAMEGAM, A. (2013), SARNO, R; SINAGA, SUNGKONO, K. R. (2020)
Manutenção (máquinas, serviços)	2	BOSE, R. P. J. C.; VAN DER AALST, W. M. P. (2013), FAZZINGA, B. et al. (2020)
Auditoria	1	HUANG, Z.; CONG, Q.; HU, J. (2012)
Não aplicável	4	JALALI, H.; BARAANI, A. (2010), VAN DER AALST, W. M. P. (2014), GENGA, L. et al. (2017), SILVA, F. A. R. (2018)

Fonte: Elaborado pelo autor (2021).

O Quadro 6 apresenta os algoritmos empregados nos estudos. Alguns estudos aplicaram mais de um tipo de algoritmo de forma comparativa, sendo verificado a utilização híbrida da mineração de processos com métodos de mineração de dados como: algoritmos de aprendizado de máquina, regras de associação, agrupamento e árvores de decisão.

QUADRO 6 – RESULTADO DA REVISÃO DA LITERATURA

Algoritmo	Autor(es)
Algoritmo alpha	VAN DER AALST, W. M. P.; DE MEDEIROS, A. K. A. (2005), HUANG, Z.; CONG, Q.; HU, J. (2012), BEZERRA, F.; WAINER, J. (2013)
Algoritmo alpha ++	HUANG, Z.; CONG, Q.; HU, J. (2012), BEZERRA, F.; WAINER, J. (2013), CAESARITA, Y.; SARNO, R.; SUNGKONO, K. R. (2017), RAHMAWATI, D.; SARNO, R. (2017)
Algoritmo genético	HUANG, Z.; CONG, Q.; HU, J. (2012), CHAO, O.; CHENG, H.; JUAN, Y. (2014)
Minerador heurístico	MARDANI, S.; SHAHRIARI, H. R. (2013), BEZERRA, F.; WAINER, J. (2013), CAESARITA, Y.; SARNO, R.; SUNGKONO, K. R. (2017), RAHMAWATI, D.; YAQIN, M. A.; SARNO, R. (2016)
Indutive miner	TAVARES, G. M. <i>et al.</i> (2018)
Algoritmo novo	YANG, W.-S.; HWANG, S.-Y. (2006), BEZERRA, F.; WAINER, J. (2011)
Fuzzy miner (minerador difuso)	JANS, M. <i>et al.</i> (2011), PANE, E. S.; VIBAWA, A. D.; PURNOMO, M. H. (2017), FOLINO, F.; GUARASCIO, M.; PONTIERI, L. (2017), BAADER, G.; KRCCMAR, H. (2018), ZERBINO, P. <i>et al.</i> (2018),
Aprendizado de máquina (classificação, decisão, regras)	SARNO, R. <i>et al.</i> (2015), PAULRAJ, P.; NEELAMEGAM, A. (2013), CUZZOCREA, A. <i>et al.</i> (2017), FOLINO, F.; GUARASCIO, M.; PONTIERI, L. (2017), RAHMAWATI, D.; SARNO, R. (2017), CUZZOCREA, A. <i>et al.</i> (2018), TAVARES, G. M. <i>et al.</i> (2018), NOLLE, T. <i>et al.</i> (2018), SARNO, R.; SINAGA, SUNGKONO, K. R. (2020)
Métodos estatísticos, Markov, Monte Carlo	RAHMAWATI, D. <i>et al.</i> (2017), NOLLE, T. <i>et al.</i> (2018), FAZZINGA, B. <i>et al.</i> (2020)
Sem indicação de algoritmo	BOSE, R. P. J. C.; AALST, W. M. P. (2013), BOSE, R. P. J. C.; VAN DER AALST, W. M. P. (2013), VAN DER AALST, W. M. P. (2014), ADRIANSYAH, A. <i>et al.</i> (2014), HUDA, S. <i>et al.</i> (2014), CHIU, T.; JANS, M. (2017), GENGA, L. <i>et al.</i> (2017), SILVA, F. A. R. (2018)

Fonte: Elaborado pelo autor (2021).

Os algoritmos genéticos, heurísticos e difuso foram encontrados na leitura dos artigos, além dos métodos estatísticos e dos algoritmos nativos da mineração de processos como o Alpha e o Alpha++. Dentre artigos os cinco mais referenciados conforme a análise bibliométrica constam na Tabela 3.

O primeiro artigo na listagem foi publicado em 2006, abordando a identificação de fraudes em procedimentos clínicos, e a substituição das análises manuais por sistemas automatizados; a pesquisa utilizou-se de dados reais do Programa Nacional de Seguro de Saúde, de Taiwan e não empregou ferramentas específicas de mineração de processos (YANG, HWANG, 2006). Outros artigos apresentados no

Quadro 6 também não utilizaram ferramentas próprias de mineração de processos, como Bezerra, Wainer (2011) que focaram no desenvolvimento de um novo algoritmo; Pauraj, Neelamegam (2013) com o emprego de técnicas de aprendizado de máquina (*clustering*).

Ainda não foram utilizadas ferramentas específicas de mineração de processos nos estudos de Huda et al. (2014), Cuzzocrea et al. (2017), Rahmawati et al. (2017), Cuzzocrea et al. (2018), Tavares et al. (2018), Nolle et al. (2018), Fazzinga et al. (2020).

TABELA 3 – CINCO ARTIGOS MAIS REFERENCIADOS

Nome do Artigo	Número de citações
"A process-mining framework for the detection of healthcare fraud and abuse" (YANG, HWANG, 2006)	131
"Process mining and security: Detecting anomalous process executions and checking process conformance" (VAN DER AALST, DE MEDEIROS, 2005)	84
"A business process mining application for internal transaction fraud mitigation" (JANS, VANHOOF, 2011)	77
"Measuring precision of modeled behavior" (ADRIANSYAH et al., 2014)	54
"Hybrid association rule learning and process mining for fraud detection" (SARNO et al., 2015)	43

Fonte: Fujiwara, Tsunoda (2020).

O professor Willibrordus M. P. Van der Aalst (RWTH Aachen University da Alemanha) é um dos pesquisadores mais referenciados no âmbito da temática de mineração de processos e fraude conforme estudo de Fujiwara e Tsunoda (2020). O artigo "Process mining and security: detecting anomalous process executions and checking process conformance" em coautoria com a pesquisadora Ana Karla Alves de Medeiros aborda o uso de técnicas de mineração de processos para a identificação de padrões ou sua ausência nas inspeções em trilhas de auditorias, com utilização da ferramenta EMiT – Enhanced Mining Tool em um conjunto de dados de comércio eletrônico (VAN DER AALST, DE MEDEIROS, 2005).

A pesquisa de Jans et al. (2011) aplica a mineração de processos com o emprego da ferramenta ProM analisando as aquisições de bens e serviços (ciclo de compras), que envolve desde a criação do pedido, liberação, entrada do produto, faturamento chegando ao pagamento. Estudos como os de Rahmawati, Yaqin e Sarno (2016) e de Baader e Krcmar (2018) também foram aplicados em processos de aquisições, com a utilização das ferramentas ProM Process Mining e do Celonis Process Mining, respectivamente.

“Measuring precision of modeled behavior”, neste estudo os pesquisadores propõem um método para verificação da precisão dos modelos de processos, em diversos registros de eventos, como processos hospitalares, tratamento de pacientes e solicitação de alvarás de construção e o emprego da ferramenta ProM e uma extensão denominada CoBeFra⁴, um acrônimo de Comprehensive Benchmarking Framework (ADRIANSYAH et al., 2014).

A aplicação conjunta de métodos de mineração de dados e mineração de processos é percebida em diversos estudos, como em Sarno et al. (2015) onde os pesquisadores aplicam métodos de regras de associação associado à métodos de mineração de processos, aplicando no ProM sobre registros de eventos de operações com cartão de crédito objetivando a identificação de fraudes.

Nos estudos de Pauraj e Neelamegam (2013) e de Tavares et al. (2018) utiliza-se o aprendizado de máquinas com a técnica de clusterização; Cuzzocrea et al. (2017), Rahmawati e Sarno (2017), Cuzzocrea et al. (2018) aplicam árvores de decisão; Folino, Guarascio e Pontieri (2017) aplicam *fuzzy miner*, clusterização e regras de associação; as regras de associação também são identificadas no estudo de Sarno, Sinaga e Sungkono (2020).

Percebe-se que o emprego da mineração de dados é mais evidente entre os estudos relacionados a identificação de fraudes (APÊNDICE A) em relação aos de mineração de processos.

⁴ Disponível em: <http://processmining.be/cobefra/>. Permite a importação de arquivos nos formatos XES ou MXML e utiliza as redes de Petri como modelo de processo; permite a comparação e repetição de métricas de conformidade.

4.2 COMPARATIVO ENTRE FERRAMENTAS






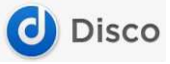


Dos levantamentos realizados nos trabalhos relacionados (comparação de ferramentas), nos estudos envolvendo a mineração de processos e a fraude, além da pesquisa na internet, selecionaram-se 8 ferramentas de mineração de processos para serem analisados no aspecto de importação e exportação de registros, e de acordo com as técnicas de mineração de processos. Soma-se a inclusão das especificações gerais da ferramenta como a versão, o código, plataforma de execução e conexão com banco de dados.

O Quadro 7 relaciona as ferramentas de mineração de processos analisadas, descrevendo a página da internet onde está disponível para download da versão *free*, de teste ou acadêmica, bem como as informações das versões comerciais para aquisição. Também consta a versão analisada, o tipo de licença ou código-fonte, a plataforma de execução, o sistema operacional quando a versão for para instalação local, a conexão com banco de dados e o tipo de ferramenta, se dedicado ou embutido.

Quanto a conexão com os bancos de dados as ferramentas suportam drivers do tipo ODBC – *Open Database Connectivity* ou do tipo JDBC – *Java Database Connectivity*, ou ambos, o Apromore não oferece suporte à conexão com banco de dados, e a ferramenta Disco disponibiliza uma aplicação (Airlift API) que permite a conexão com a fonte de dados. Em termos de ser uma ferramenta dedicada, isto é, se foi desenvolvido especificamente para mineração de processos, como é o caso do ProM, Celonis, Disco, Aris Process Mining e o Apromore; o PM4PY também é uma solução específica para mineração de processos desenvolvida na linguagem Python, o bupaR é um pacote da linguagem R e por fim o Knime, uma ferramenta de análise para ciência de dados, que possui extensões para aplicação de mineração de processos.

O segundo bloco de informações (Quadro 8) refere-se ao formato de arquivo que a ferramenta suporta, tanto para importação quanto para exportação. Todas as opções analisadas suportam a importação de arquivos do tipo CSV – *Comma-separated values*, que contém valores separados por vírgula.

QUADRO 7 – ESPECIFICAÇÃO DA FERRAMENTA DE MINERAÇÃO DE PROCESSOS

Ferramenta	ProM	Knime	bupaR	PM4Py	Celonis	Disco	Aris	Apromore
								
Página da ferramenta	promtools.org	knime.com	bupar.net	pm4py.fit.fraunhofer.de/	celonis.com	fluxicon.com	aris-process-mining.com	apromore.org
Versão analisada	Prom 6.10 with 64-bit JRE8	KNIME Analytics Platform Versão 4.2.2	Versão 0.4.2	Release 2.2.1	Celonis Intelligent Business Cloud	Versão 2.11.2	Versão 10	Enterprise Edition 7.16
Licença / código	Código aberto, GNU General Public License	Código aberto, GNU General Public License	Código aberto	Código aberto	Livre, acadêmico, comercial	Acadêmica, comercial	Livre, acadêmico, comercial	Livre, acadêmico, comercial
Plataforma de execução	Instalação local	Instalação local	Instalação local	Instalação local	Nuvem, instalação local, híbrido	Instalação local	Software como serviço (SaaS), instalação local, nuvem	Software como serviço, local (SaaS). Instalação local
Sistema Operacional	Mac OS X, Windows, Linux	Windows, Linux, Mac OS 10.13 ou superior	Windows, Linux e MacOS	Windows, Linux e MacOS		macOS 10.10 ou superior Windows 7 SP1 (x64) ou superior		Windows, Linux e MacOS
Conexão com banco de dados	ODBC drivers, JDBC drivers	JDBC drivers	ODBC drivers, JDBC drivers	ODBC drivers	JDBC drivers	Airlift API	JDBC drivers	Não disponível
Tipo de ferramenta (dedicado/embutido)	Dedicado	Embutido (plugin)	Embutido (pacote do R)	Dedicado (desenvolvido em python)	Dedicado	Dedicado	Dedicado	Dedicado

Fonte: Elaborado pelo autor (2021).

QUADRO 8 – CARACTERÍSTICAS DE IMPORTAÇÃO E EXPORTAÇÃO DOS REGISTROS DE EVENTOS

Ferramenta		ProM	Knime	bupaR	PM4Py	Celonis	Disco	Aris	Apromore
Formato de arquivo de importação suportado	CSV	✓	✓	✓	✓	✓	✓	✓	✓
	XLS					✓	✓	✓	
	XLSX						✓		
	MXML	✓	✓				✓		✓
	MXML.GZ		✓				✓		✓
	XES	✓	✓	✓	✓	✓	✓		✓
	XES.GZ		✓				✓		✓
	FXL						✓		
	DSC						✓		
Formato de arquivo de exportação suportado	CSV	✓	✓		✓	✓	✓	✓	✓
	XLS								
	XLSX					✓			
	MXML	✓					✓		
	MXML.GZ	✓					✓		
	XES	✓		✓	✓	✓	✓		
	XES.GZ	✓					✓		
	FXL	✓					✓		
	DSC						✓		
	PNG	✓					✓	✓	✓
	JPEG	✓					✓	✓	
	PDF	✓					✓	✓	✓

Fonte: Elaborado pelo autor (2021).

QUADRO 9 – SUPORTE AS TÉCNICAS DE MINERAÇÃO DE PROCESSOS

Ferramenta		ProM	Knime	bupaR	PM4Py		Celonis	Disco	Aris	Apromore
Modelo de visualização do processo	Redes de Petri	✓	✓	✓	✓					
	BPMN	✓			✓		✓		✓	✓
	Modelos <i>fuzzy</i>	✓						✓		
	Modelos heurísticos	✓	✓							
	Indutive Miner	✓	✓							
Anonimização de dados							✓	✓	✓	
Descoberta de processos		✓	✓	✓	✓		✓	✓	✓	✓
Verificação de conformidade		✓	✓	✓	✓		✓	✓	✓	✓
Aprimoramento do processo							✓	✓		
Aplicação de filtros		✓	✓		✓		✓	✓	✓	✓
Análise de redes organizacionais (sociais)		✓		✓	✓		✓			
Regras de decisão		✓	✓		✓					
Análise de performance		✓	✓				✓	✓	✓	✓
Simulação					✓					

Fonte: Elaborado pelo autor (2021).

O último bloco de características contempla a identificação das técnicas de mineração de processos suportada pelas ferramentas analisadas. Conforme Quadro 9 todas as opções permitem a descoberta de processos e a verificação de conformidade. As ferramentas ProM, bupaR, PM4PY e Celonis permitem a análise de redes sociais ou redes organizacionais. Com base nestes levantamentos.

Os modelos de processos em algumas ferramentas permitem uma análise visual de gargalos, de inconsistências tanto em questão de sequência de atividades ou quanto ao tempo de execução ou retrabalho. Ferramentas como o Disco possibilitam uma interação rápida de simplificação dos modelos, através de comandos de redução de atividades ou dos caminhos entre as atividades. No Apêndice B nota-se a importância deste recurso para uma análise de tomada de decisão.

A diferenciação entre versões de ferramentas de mineração de processos, além da mudança visual, pode apresentar a inclusão ou exclusão de componentes, melhorias ou piora no desempenho nos cálculos e aplicações de algoritmos. No Apêndice C um exemplo de mudança visual entre as versões do ProM da 5.2 para 6.1.

Na subseção foram utilizadas três das ferramentas apresentadas para serem analisadas, a primeira o ProM que aparece em grande parte dos estudos levantados nas subseções anteriores, o Disco da Fluxicon que disponibiliza uma versão acadêmica que pode ser instalada no computador e o Celonis Snap, uma versão gratuita com utilização em nuvem, esta versão é diferente da analisada no mapa comparativo, mas que permite a descoberta de processos e sua visualização.

4.3 SIMULAÇÕES NA FERRAMENTA DE MINERAÇÃO DE PROCESSOS

Nesta subseção apresenta-se a análise de um conjunto de registros disponibilizados no sítio do Tribunal de Contas do Estado do Paraná, contendo as obras executadas nos municípios do Paraná e as respectivas medições. Realizada a construção da base de dados, com auxílio de duas ferramentas de mineração de processos, pôde-se julgar o seu emprego na prevenção de fraudes.

No tratamento dos registros, emprego da mineração da mineração de processos e análise dos resultados empregaram-se os recursos:

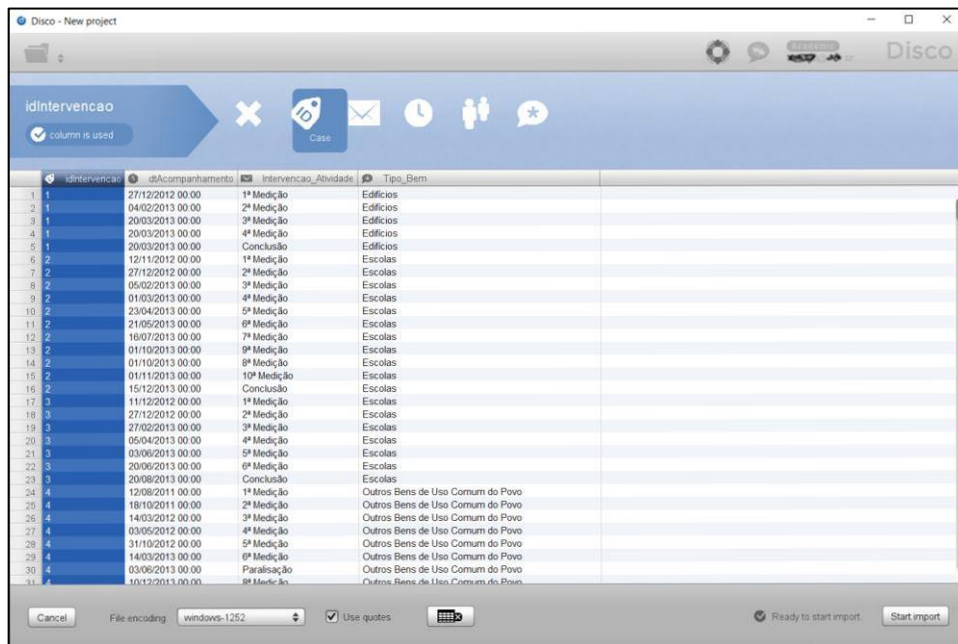
- a) Laptop com processador Intel® Core™ I5-7300 @2,50GHz; memória RAM de 16,0 GB; sistema operacional Windows 10 de 64 bits, versão 20H2; dispositivo gráfico Nvidia GeForce GTX 1050; e armazenamento SSD de 256GB, 1800MB/s de leitura e 900 MB/s de gravação;
- b) Microsoft Excel 365 versão 2101;
- c) ProM 6.10 revisão 43952;
- d) Disco versão 2.11.2, release 09/02/2021;
- e) Celonis Snap.

A extração e seleção dos dados resultaram em 325.749 registros preliminares, onde o primeiro registro tem como data de intervenção “24/08/1992 00:00” e o último registro a data “31/01/2021 00:00”. O campo de detalhamento do tipo do bem resultou em 59 variantes e de registros sem identificação do tipo de bem relacionado; o conjunto contém 35.990 intervenções, cada intervenção refere-se a uma obra executada que foi relacionada de forma generalizada sem a identificação do município que estão localizados ou características físicas.

Considerando a informação do TCE-PR, o espectro temporal dos registros é de janeiro de 2003 até a última atualização (16/02/2021 neste estudo), verificou-se a existência de registros de intervenção anteriores a esta data. Na primeira limpeza excluíram-se os registros sem o devido relacionamento com o tipo de bem, equivalente a 2.939 registros e 458 intervenções, e as 139 intervenções com data anterior a 01/01/2013, referente a 356 registros.

O conjunto final para análise resultou em 322.453 registros ou 98,99% dos dados brutos, exportados em um arquivo do tipo CSV com a utilização do software Microsoft Excel. A Figura 16 apresenta a tela da ferramenta Disco com as primeiras configurações para importação do conjunto de registros. No estudo seguiu-se o relacionamento do “Case” com o campo “IdIntervencao”, “Timestamp” com o campo “dtAcompanhamento”, “Activity” com o campo “Intervencao_Atividade” e “Other” com o campo “Tipo_Bem”.

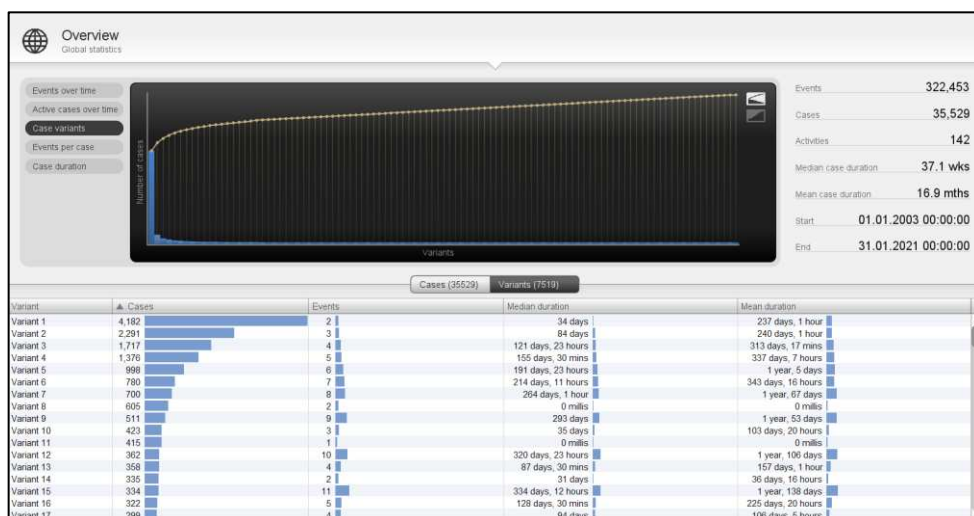
FIGURA 16 – IMPORTAÇÃO DO ARQUIVO CSV NA FERRAMENTA DISCO



FONTE: Elaborada pelo autor (2021).

Os 322.453 registros (*events*) referem-se a 35.529 intervenções (*cases*) e 142 atividades que incluem medições, conclusões, paralizações e cancelamentos de intervenções. A Figura 17 apresenta a composição do conjunto de registros sem a distinção do tipo de obra executada com 7.519 variações de casos, compreendendo o período de 01/01/2003 a 31/01/2021. A 1ª variante compreende 4.182 casos com 2 eventos, uma duração mediana de 34 dias e média de 237 dias e 1 hora.

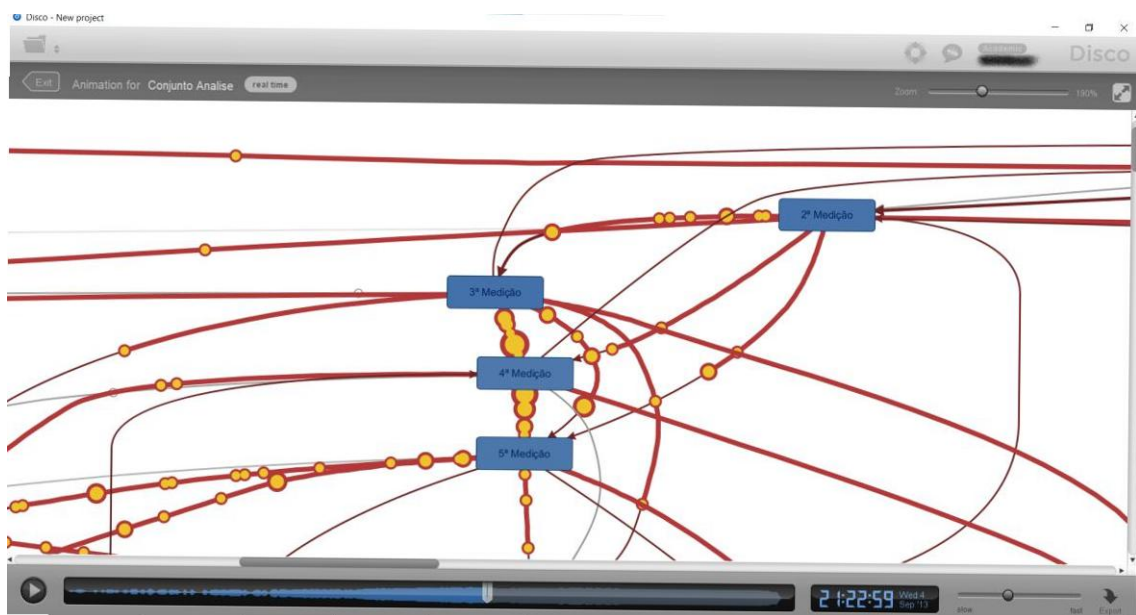
FIGURA 17 – RETRATO DO CONJUNTO DE REGISTROS



FONTE: Elaborada pelo autor (2021).

Com aplicação de um filtro no atributo “tipo_bem”, selecionou-se para análise somente as obras que envolvam as escolas, que representam 2.157 casos (6,07% do total) e 23.061 eventos (7,15% do total). A ferramenta Disco possui alguns recursos que permitem a visualização temporal da execução dos eventos e a identificação de atividades com maior de execução, possíveis gargalos ou indicação de atividade ignorada. Na Figura 18, um exemplo onde da 2ª medição pula para a 4ª ou 5ª medição.

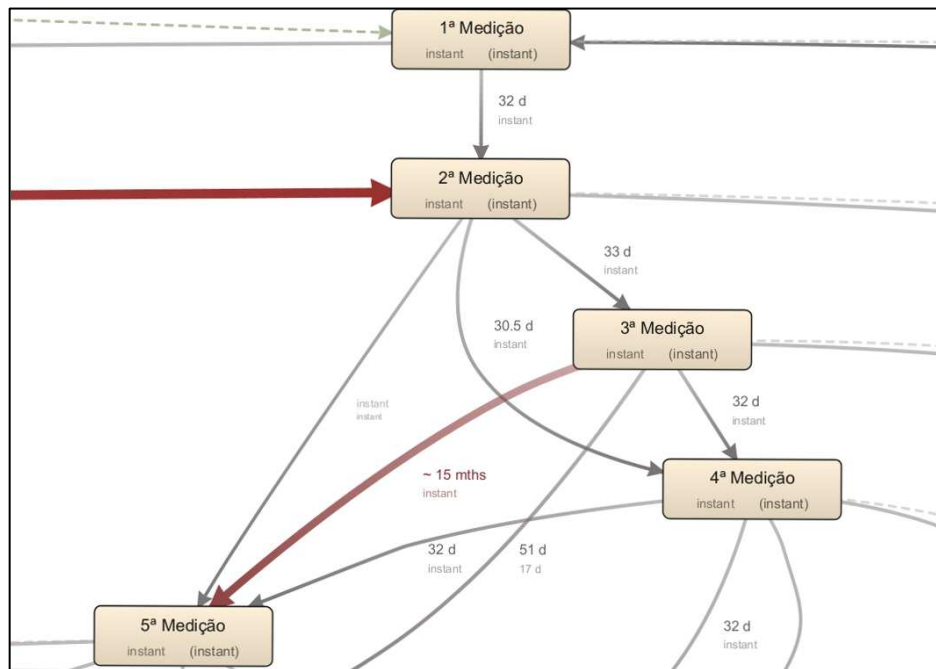
FIGURA 18 – TELA DE ANIMAÇÃO DO FLUXO DOS EVENTOS DA FERRAMENTA DISCO



FONTE: Elaborada pelo autor (2021).

A análise gráfica do tempo mediano das execuções das atividades é um dos recursos que a ferramenta Disco disponibiliza, que possibilita uma análise rápida como se verifica na Figura 19. No exemplo, a duração mediana entre a 1ª medição e a 2ª medição é de 32 dias, da 2ª para 3ª medição de 33 dias, e da 3ª para 4ª medição de 32 dias, que indica um padrão de recorrência entre as medições. As ocorrências divergentes indicam circunstâncias que requerem uma análise mais aprofundada, ainda no mesmo exemplo verificamos casos em que o tempo decorrido entre as medições é instantâneo.

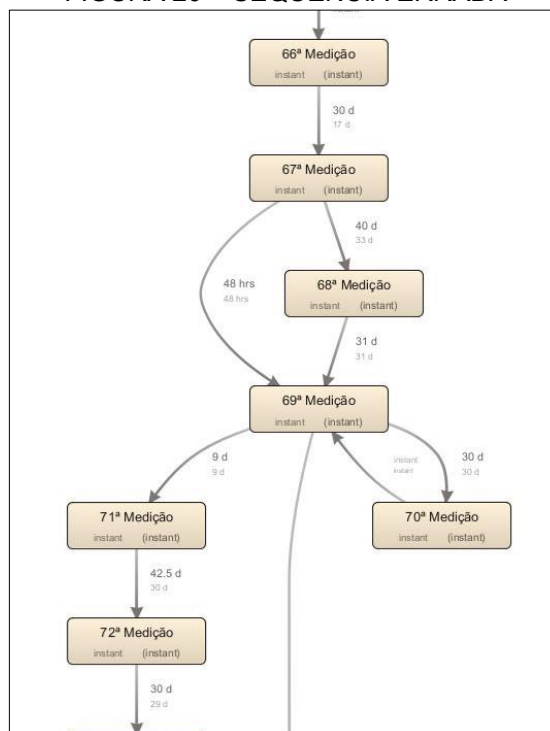
FIGURA 19 – RECORTE DO TEMPO MEDIANO DAS ATIVIDADES



FONTE: Elaborada pelo autor (2021).

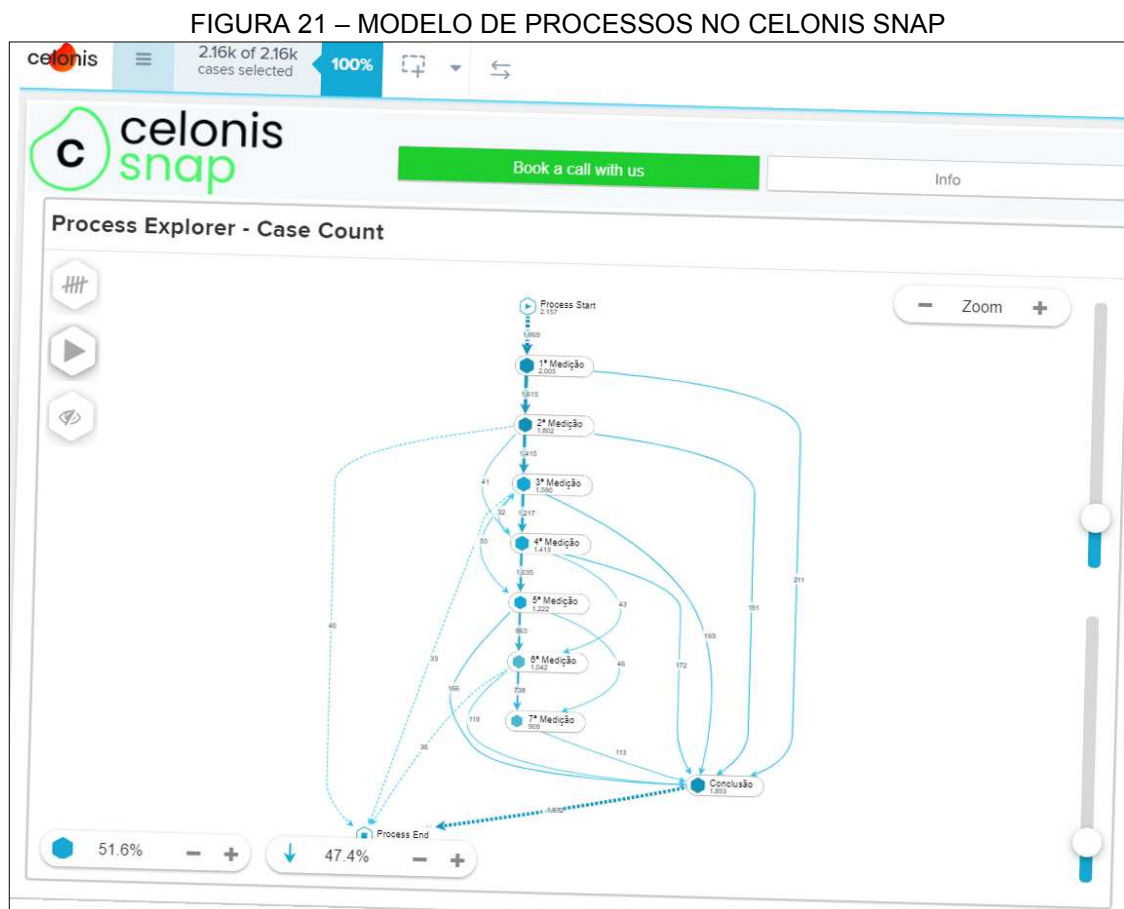
Outro exemplo de sequência errada na visualização de modelo de processo na ferramenta Disco, onde se identifica que da 67ª medição o fluxo de trabalho pula para 69ª medição (FIGURA 20).

FIGURA 20 – SEQUÊNCIA ERRADA



FONTE: Elaborada pelo autor (2021).

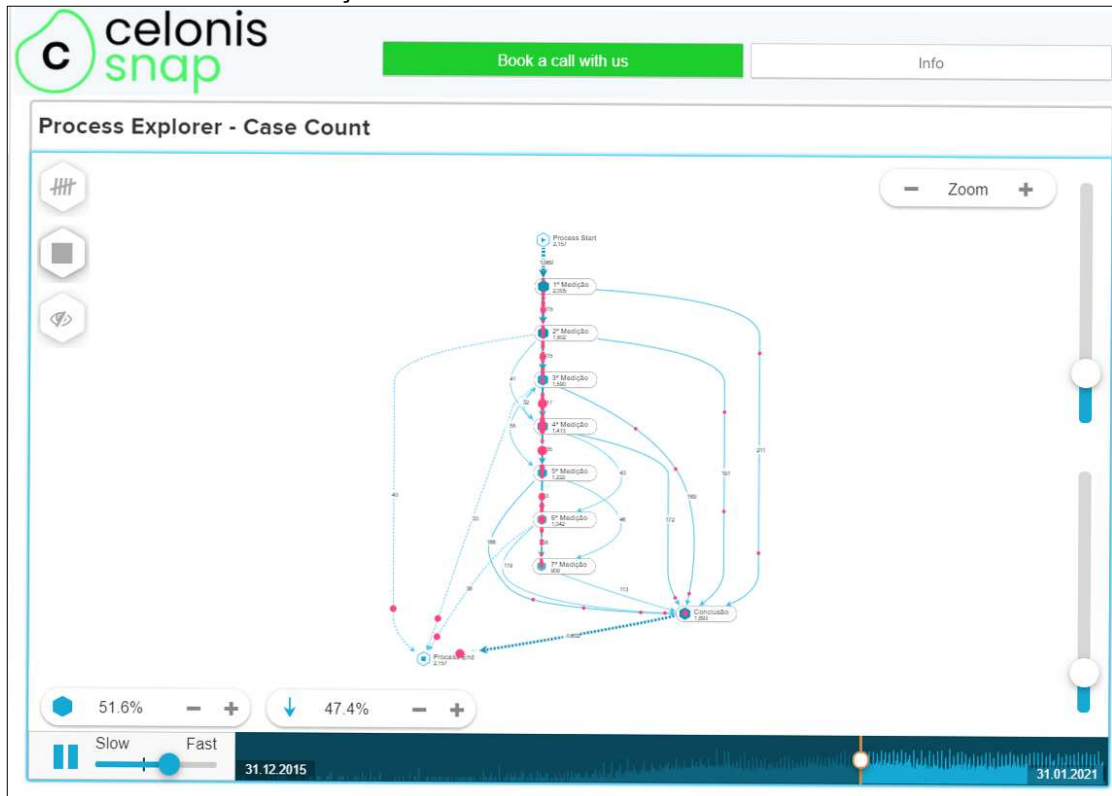
Com a utilização do mesmo conjunto de registros no Celonis Snap, diferentemente do Disco a execução desta ferramenta é na nuvem, também possui uma interface visual amigável para análise do processo, que permite identificar casos em que a sequência das atividades está diferente do padrão. Recursos de simplificação do modelo e dos caminhos das atividades possibilitam identificação e triagem dos casos para uma auditoria documental ou presencial no local da execução da obra (FIGURA 21).



FONTE: Elaborada pelo autor (2021).

O recurso de animação do fluxo de atividades do processo, de acordo com o tempo de execução de atividades visto no Disco também está presente no Celonis. Na Figura 22 identifica-se o intervalo de tempo dos registros compreendendo o período de 31/12/2015 a 31/01/2021, os pontos vermelhos indicam o percurso de cada caso e permite, de acordo com a intensidade, a percepção de gargalos temporais.

FIGURA 22 – ANIMAÇÃO DO MODELO DE PROCESSOS NO CELONIS SNAP

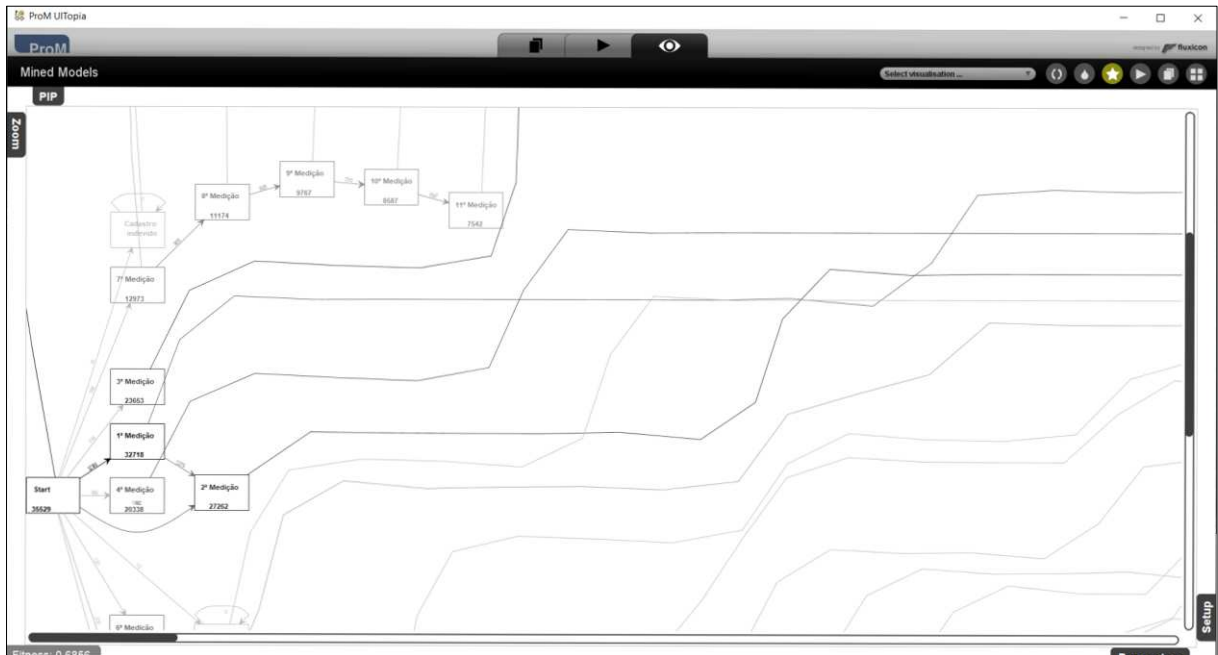


FONTE: Elaborada pelo autor (2021).

A duas ferramentas apresentadas (Disco e Celonis) possuem uma interface de fácil compreensão e utilização para usuário, com um conjunto de registros de eventos formado basicamente por um identificado de caso ou evento, compostos por atividades vinculadas com uma data ou data e hora, já possibilita a importação e análise nas ferramentas. Nestas simulações não foram empregados, mas as ferramentas permitem, a identificação de outras características e recursos (equipamentos ou funcionários) que possibilitam a realização de outras análises como de redes sociais na organização.

A ferramenta ProM, por ser de código aberto, possui um rol maior de algoritmos e funções que foram adicionadas ao longo dos anos pelos pesquisadores. Tanto os estudos envolvendo a mineração de processos com a identificação de fraudes como os estudo de comparação de ferramentas, em geral, utilizam ou citam esta ferramenta. Na Figura 23 apresenta um modelo de processos, utilizando a mesma composição de registros das simulações anteriores, porém diferenciando pela sequência de atividades estarem no sentido horizontal da direita para esquerda, e não na vertical de cima para baixo nas simulações anteriores.

FIGURA 23 – MODELO DE PROCESSO NA FERRAMENTA PROM



FONTE: Elaborada pelo autor (2021).

A ferramenta ProM visualmente tem um aspecto menos amigável para o usuário, requerendo um conhecimento mais técnico para sua utilização. Na percepção do pesquisador, esta diferenciação deve-se a condição de código-aberto em que cada pesquisador pode adicionar novos recursos, testar, incrementar ou alterar algoritmos ou modelos existentes, tendo assim uma finalidade mais acadêmica do que comercial, ao contrário do desenvolvimento das ferramentas Disco e Celonis que foram criadas para facilitar a sua utilização na tomada de decisão.

As simulações foram executadas sem a descoberta prévia de um modelo de processo, porém os recursos disponíveis nas três ferramentas além de identificarem o processo padrão também permitem a identificação dos casos que fogem tanto da ordenação sequencial, quanto do tempo mediano de execução das atividades. O presente estudo a abordagem foi focada na execução de escolas seguindo a sequência numeral da medição, ou seja, 1ª medição, 2ª medição, 3ª medição, e assim sucessivamente.

Sarno, Sinaga e Sungkono (2020) listam alguns tipos de análise que são indicadores de tipologias anômalas, dentre esses temos a análise de fluxo de controle e a análise de tempo de processamento, que foram aplicados no estudo. Quanto ao fluxo de controle a análise remete a dois atributos se uma atividade foi ignorada ou executada em desacordo com um padrão estabelecido. Ao verificar no exemplo da Figura 20 que da 67ª medição o fluxo de trabalho pula para 69ª medição, ignorando

uma a 68ª medição o analista prontamente pode identificar que ocorreu uma anomalia no processo, sinalizando um ponto a ser averiguado solicitando maiores informações ou realizando inspeções.

Do ponto de vista do tempo de processamento a análise dos fluxos apresenta um tempo mediano entre as medições de 32 dias, coerente os trâmites administrativos em órgãos governamentais. Este valor indica um padrão, um valor abaixo deste padrão é um indicador de anomalia em questão de tempo de processamento, da mesma forma valores superiores. Estes indicadores podem não refletir necessariamente fraudes, mas problemas procedimentais na própria execução da obra, necessidade de correção nos serviços executados ou serviços não executados no planejamento do cronograma.

5 CONSIDERAÇÕES FINAIS

Retornando as premissas iniciais, o estudo de Van der Aalst (2016) apontava a mineração de processos que possibilita a revelação de desvios e defeitos nos fluxos de processos. Da leitura dos artigos verifica-se a aplicação foca em áreas como na verificação de inconformidades nos hospitalares e nos tratamentos clínicos, permitindo a modelagem dos fluxos, a readequação e otimização. Nesta linha, a mineração tem o objetivo de minimizar os riscos nos processos.

Na aplicação no setor bancário temos uma aplicação que mais se aproximou do foco deste estudo, de identificar anormalidades nos processos. Nos casos de liberação de créditos bancários o intuito é verificar a conformidade nos processos, criando alertas a possíveis fraudes que possam acontecer. Da mesma forma, nas simulações realizadas nos registros de eventos TCE-PR podemos visualizar indicativos de situações anômalas, de acordo com análise de alguns atributos como a condição de execução sequencial das atividades e do tempo mediano que é necessário para executar tais atividades.

Outro ponto levantado inicialmente é a dificuldade na modelagem de algoritmos de prevenção e identificação de fraudes considerando a análise de dados desequilibrados (WANG; LI; ZHAO, 2019). No conjunto de registros de eventos analisados e posteriormente importados nas ferramentas de mineração de processos, a dificuldade não foi percebida, a questão dos dados desequilibrados parte do princípio de que num conjunto dados somente uma parte muito pequena pode ter algum indício de fraude e uma parte muito menor desta amostragem pode ser caracterizada como fraude.

Todas as três ferramentas (ProM, Disco e Celonis) permitiram a identificação de casos com atividades anômalas, que possibilitam a sua utilização como alerta e indicador para equipes de auditorias.

Outra premissa identificada é uma possível dificuldade no entendimento de algoritmos e a aplicação das técnicas de mineração de processos, justificada por desconhecimento técnico e a uma dita complexidade (BURRELL, 2016). No presente estudo, apesar da leitura de alguns estudos que tiveram um foco maior no desenvolvimento de algoritmos não ocorreu um aprofundamento na análise dos algoritmos de mineração, e sim na aplicação e manuseio das ferramentas existentes.

Esta análise demonstrou que a utilização das três ferramentas pode ser utilizada sem um conhecimento prévio aprofundado de técnicas de programação ou a necessidade desenvolvimento de algoritmos. As ferramentas permitem de forma fácil importar arquivos para serem analisados, e fornecem resultados que permitem ao analista tirar informações dos registros de eventos, talvez a maior dificuldade é na obtenção das informações formatadas ou que permitam a sua formatação para a análise.

Na trajetória de pesquisa proposta a questão norteadora apresentada foi de **Como a mineração de processos pode ser auxiliar na mitigação e prevenção de fraudes na gestão pública?** Assim, foram desenvolvidos em três etapas esta pesquisa, que se somam para encontrar a resposta do questionamento.

5.1 REVISÃO SISTEMÁTICA DA LITERATURA

A revisão de literatura envolvendo a mineração de processos e fraudes não retornou um volume expressivo de estudos, sendo que se identificou que emprego da ferramenta ProM utilizado por muitos dos pesquisadores. A sua maior utilização pode ser justificada pelo acesso livre e de código aberto, o que permite o desenvolvimento de novos recursos e algoritmos, esta ferramenta possibilita ao pesquisador testar e verificar a consistência das suas pesquisas.

Também vale ressaltar vários estudos que objetivaram a criação de novos algoritmos, o emprego conjugado da mineração de processos com a mineração de dados, principalmente no emprego de técnicas de agrupamento, classificação e árvores de decisão. Os estudos dos fluxos de atividades, recaem sobre autores referenciais como o professor Van der Aalst que em coautoria com vários pesquisadores, suportam novos estudos, inclusive o presente estudo.

Do ponto de vista dos resultados, os artigos citados na pesquisa demonstram a condição da mineração de processo como ferramenta de prevenção e identificação de fraudes, quando voltamos nossos olhares para o setor bancário. Situação que mais se assemelha ao presente estudo.

As aplicações em áreas como a saúde, também pode ser empregadas na administração pública na modelagem procedimentos, verificação de gargalos em termos servidores sobrecarregados, ou identificação na demora das análises pelos setores.

5.2 COMPARATIVO DAS FERRAMENTAS DE MINERAÇÃO DE PROCESSOS

Apontou-se 10 estudos onde os pesquisadores comparavam ferramentas de mineração de processos, sendo o primeiro publicado em 2011 e o último em 2021. Várias ferramentas foram analisadas, com diferentes versões que se alteravam durante os anos, pontos de vistas e características das mais diversas possíveis. Diante das opções existentes no mercado, selecionou-se 8 delas, aquelas de livre acesso, como versões acadêmicas ou de teste, ou também as de código-aberto, assim a análise abrangeu as ferramentas: ProM, Knime, bupaR, PM4PY, Celonis, Disco, Aris e Apromore.

A plataforma de execução empregada por estas soluções, em sua maioria, é para instalação local, no computador, porém algumas das recentes ferramentas já são disponibilizadas para utilização na nuvem. A possibilidade de integração com banco de dados, é um recurso que permite a atualização de dados sem a necessidade de importação ou exportação manual dos casos a serem analisados. Dentre as técnicas de mineração de processos propriamente ditas, a descoberta de modelos de processos e a verificação da conformidade, isto é, a constatação de que os novos casos seguem o padrão do modelo, estão presentes nas ferramentas analisadas.

Outros recursos como a anonimização de dados, aprimoramento de processos, análise de redes sociais, análise de desempenho e possibilidade de simulações de cenários foram identificados, em algumas das ferramentas. As opções destas ferramentas auxiliam do emprego da mineração de processos, deste a modelagem, visualização e da correção dos modelos existentes.

5.3 O EMPREGO DA MINERAÇÃO DE PROCESSOS EM UM CONJUNTO DE REGISTROS DE LICITAÇÕES DE OBRAS PÚBLICAS DE MUNICÍPIOS PARANAENSES.

Neste terceiro e último objetivo específico, agrupou-se os conhecimentos adquiridos nas etapas anteriores da pesquisa. Aplicando-se num conjunto de registros disponibilizados nos dados abertos do Tribunal de Contas do Estado do Paraná, contendo as medições das obras executadas pelas prefeituras municipais, porém efetuando-se um tratamento nos registros de eventos, desconsiderando-se os

recursos (nomes dos servidores) zelando-se pela anonimização dos dados, visto que o intuito foi a simulação e julgamento das ferramentas.

Vislumbrou-se que apesar da utilização maciça do ProM nos estudos relatados as ferramentas Disco e Celonis Snap, são operação mais amigável ao usuário. Todas as ferramentas permitem a identificação de ações que alertam uma inconsistência no fluxo de trabalho, como percebe-se que existem casos ondem as medições pulam o sequenciamento ou que existem o retorno as atividades anteriores.

No procedimento de identificação de possíveis fraudes na gestão pública constata-se que a ferramenta de mineração de processos demonstrou que é possível a sua aplicação. Os resultados obtidos nas simulações permitiram a identificação de situações anômalas que permitem ao analista filtram as situações que merecem maior atenção.

Considerando a impossibilidade de auditar e verificar a totalidade dos registros e das situações, a utilização que permitam selecionar uma amostra para auditoria é de extrema utilidade e que proporciona um ganho para administração pública.

Em outros aspectos, a aplicação da mineração pode ser útil na administração pública, na experiência do pesquisador no setor público, pode-se constatar a utilidade no mapeamento de processos, no controle de fluxos dos trabalhos, e identificação de gargalos.

No presente estudo, também não foram considerados os recursos empregados, uma análise com este foco pode indicar a desnecessidade de tramitação para determinar setor ou a inversão de ordem dos fluxos. Também a identificar servidores que estão centralizando as atividades, sendo sobrecarregados ou pouco utilizados.

No controle de tempo de resposta ao contribuinte, como na solicitação de serviços, de informações ou atendimento a reclamações, pode-se identificar o tempo mínimo, médio e máximo da execução deles, com isso informar previamente o tempo de atendimento da solicitação.

5.4 SUGESTÕES PARA CONTINUIDADE DO ESTUDO

As pesquisas limitam-se e são delimitadas num campo específico de abrangência, por motivos do escasso tempo para o seu desenvolvimento e o oceano

de oportunidades que são vislumbradas no percurso do conhecimento. Nos encaminhamentos desta pesquisa – de muitas leituras – percebe-se que a mineração de processos pode ser empregada em diversos fluxos de processos, no âmbito da administração pública, como em trâmites de protocolos, solicitações de licenças urbanísticas, autorizações comerciais, e no âmbito das empresas privadas, como citados em alguns estudos, em fluxos de aquisições de bens e serviços, protocolos hospitalares, permitindo a identificação de oportunidades e ganhos para empresa.

Indaga-se os benefícios na aplicação em outras áreas do conhecimento, fora aqueles já relatados nos estudos, como ferramental na governança corporativa, nas auditorias de *compliance*, na transparência do *accountability*. Além disso, nos trabalhos futuros sugere-se uma análise de outras ferramentas, não apenas com os dados apresentados em estudos ou documentos de fornecedores, mas uma visão de usuário e gestor responsável por uma tomada de decisão.

Assim, estas considerações e recomendações finalizam os alicerces deste estudo, que complementam outros, e que continuam nas futuras pesquisas.

REFERÊNCIAS

- ADRIANSYAH, A. et al. Measuring precision of modeled behavior. **Information Systems and e-Business Management**, v.13, n.1, p. 37-67, 2014. Disponível em: <https://doi.org/10.1007/s10257-014-0234-7>. Acesso em: 04 jul. 2020.
- AGARWAL, N.; SINGH, L. Process mining tools: a comparative analysis and review. **Advances in Computer Science and Information Technology (ACSIT)**, v.1, n.2., nov.2014, pp. 26-29, 2014. Disponível em: <http://www.krishisanskriti.org/acsit.html>. Acesso em: 14 fev. 2021.
- AILENEI, I-M. **Process mining tools: a comparative analysis**. Thesis (Master of Science in Business Information Systems) – Department of Mathematics and Computer Sciences, Eindhoven University of Technology, Eindhoven, 2011.
- APPOLINÁRIO, F. **Dicionário de metodologia científica: um guia para a produção do conhecimento científico**. 2 ed. São Paulo: Atlas, 2011.
- ARAÚJO, V.S.; ZULLO, B.A.; TORRES, M. Big data, algoritmos e inteligência artificial na administração pública: reflexões para a sua utilização em um ambiente democrático. **A&C – R. de Dir. Administrativo & Constitucional**, ano 20, abr./jun, n. 80, p. 241-261, Belo Horizonte, 2020. Disponível em: <http://dx.doi.org/10.21056/aec.v20i80.1219>. Acesso em: 23 mar. 2021.
- BAADER, G.; KRCCMAR, H. Reducing false positives in fraud detection: Combining the red flag approach with process mining. **International Journal of Accounting Information Systems**, v. 31, 2018. Disponível em: <http://dx.doi.org/10.1016/j.accinf.2018.03.004>. Acesso em: 04 jul. 2020.
- BAUDER, R.A.; ROSA, R.C. da; KHOSGOFTAAR, T.M. Identifying medicare provider fraud with unsupervised machine learning. **Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018**, art. no. 8424722, p. 285-292, 2018. Disponível em: <https://doi.org/10.1109/IRI.2018.00051>. Acesso em: 01 nov. 2019.
- BEVERUNGEN, D., et al. Seven paradoxes of business process management in a hyper-connected world. **Business & Information Systems Engineering**, 2020. Disponível em: <https://doi.org/10.1007/s12599-020-00646-z>. Acesso em: 22 ago. 2020.
- BEZERRA, F.; WAINER, J. Fraud detection in process aware systems. **International Journal of Business Process Integration and Management**, v.5, n.2, p. 121-129, 2011. Disponível em: <https://dx.doi.org/10.1504/IJBPIIM.2011.040204>. Acesso em: 04 jul. 2020.
- BHATTACHARYYA, S. et al. Data mining for credit card fraud: a comparative study. **Decision Support Systems**, v. 50, n.3, p. 602-613, 2011. Disponível em: <https://doi.org/10.1016/j.dss.2010.08.008>. Acesso em: 20 jun. 2020.

BIEGELMAN, M. T.; BARTOW, J. T. **Executive roadmap to fraud prevention and internal control: creating a culture of compliance**. 2 ed. John Wiley & Sons: New Jersey, 2012.

BOSE, R. P. J. C.; VAN DER AALST, W. M. P. Discovering signature patterns from event logs. **2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)**, p. 111-118, 2013. Disponível em: <https://doi.org/10.1109/CIDM.2013.6597225>. Acesso em: 04 jul. 2020.

BOSE, R.P.J.C., VAN DER AALST, W.M.P. Process Mining Applied to the BPI Challenge 2012: Divide and Conquer While Discerning Resources. In: La Rosa M., Soffer P. (eds) Business Process Management Workshops. BPM 2012. **Lecture Notes in Business Information Processing**, vol 132. Springer, Berlin, Heidelberg, 2013. Disponível em: https://doi.org/10.1007/978-3-642-36285-9_25. Acesso em: 04 jul. 2020.

BRU, F.; CLAES, J. The perceived quality of process discovery tools. **axXiv**, 2018. <https://arxiv.org/ftp/arxiv/papers/1808/1808.06475.pdf>. Acesso em: 15 fev. 2021.

BURATTIN, A. **Process mining techniques in business environments**: theoretical aspects, algorithms, techniques, and open challenges in process mining. Lecture Notes in Business Information Processing, v. 207. Cham: Springer, 2015.

BURRELL, J. How the machine 'thinks': understanding opacity in machine learning algorithms. **Big Data and Society**, v.3, n.1, 2016. Disponível em: <https://doi.org/10.1177%2F2053951715622512>. Acesso em: 19 jun. 2020.

CAESARITA, Y.; SARNO, R.; SUNGKONO, K. R. Identifying bottlenecks and fraud of business process using alpha ++ and heuristic miner algorithms (case study: CV. wicaksana artha). In *Proceedings of the 11th International Conference on Information and Communication Technology and System, ICTS 2017*, p. 143-148, 2018. Disponível em: <https://doi.org/10.1109/ICTS.2017.8265660>. Acesso em: 04 jul. 2020.

ÇELIK, U.; AKÇETIN, E. Process mining tools comparison. **AJIT-e: Online Academic Journal of Information Technology**, v.9, n.34, 2018. Disponível em: <https://doi.org/10.5824/1309-1581.2018.4.007.x>. Acesso em: 14 fev. 2021.

CHAO, O.; CHENG, H.; JUAN, Y. An integrated mining approach to discover business process models with parallel structures: towards fitness improvement. **International Journal of Production Research**, n.53, v.13, p. 3888-3916, 2014. Disponível em: <https://doi.org/10.1080/00207543.2014.974847>. Acesso em: 04 jul. 2020.

CHIU, T.; JANS, M. Process Mining of Event Logs: A Case Study Evaluating Internal Control Effectiveness. **Social Science Research Network (SSRN)**, 05. jul. 2017. Disponível em: <http://dx.doi.org/10.2139/ssrn.3136043>. Acesso em: 04 jul. 2020.

CONSTANTE-NICOLALDE, F. V.; GUERRA-TERÁN, P.; PÉREZ-MEDINA, J. L. Fraud prediction in smart supply chains using machine learning techniques. In: Botto-Tobar M., Zambrano Vizuete M., Torres-Carrión P., Montes León S., Pizarro Vásquez G., Durakovic B. (eds) Applied Technologies. ICAT 2019. **Communications in Computer**

and Information Science, v. 1194. Springer, Cham, 2020. Disponível em: https://doi.org/10.1007/978-3-030-42520-3_12. Acesso em: 27 jul. 2020.

CUZZOCREA, A. et al. Experimenting and Assessing a Probabilistic Business Process Deviance Mining Framework Based on Ensemble Learning. In: Hammoudi S., Śmiałek M., Camp O., Filipe J. (eds) Enterprise Information Systems. ICEIS 2017. **Lecture Notes in Business Information Processing**, v. 321. Springer, Cham, 2017. Disponível em: https://doi.org/10.1007/978-3-319-93375-7_6. Acesso em: 04 jul. 2020.

CUZZOCREA, A. et al. Extensions, Analysis and Experimental Assessment of a Probabilistic Ensemble-learning Framework for Detecting Deviances in Business Process Instances. *In Proceedings of the 19th International Conference on Enterprise Information Systems*, v.3: ICEIS, p. 162-173, 2017. Disponível em: <https://doi.org/10.5220/0006340001620173>. Acesso em: 04 jul. 2020.

DE SMET, D.; MAYER, N. Integration of it governance and security risk management: A systematic literature review. **International Conference on Information Society** (i-Society), Dublin, p. 143-148, 2016. Disponível em: <https://doi.org/10.1109/i-Society.2016.7854200>. Acesso em: 12 ago. 2020.

DESAI, N. Understanding the theoretical underpinnings of corporate fraud. **Vikalpa**, v. 45, n. 1, Mar. 2020, p. 25–31, 2020. Disponível em: <https://doi.org/10.1177/0256090920917789>. Acesso em: 12 ago. 2020.

DEVI, T. A.; KUMUDAVALLI, M. V.; SUDHAMANI. An informative and comparative study of process mining. **International Journal of Scientific & Engineering Research**, v.8, n.8, 2017. Disponível em: <https://www.ijser.org/onlineResearchPaperViewer.aspx?An-Informative-and-Comparative-Study-of-Process-Mining-Tools.pdf>. Acesso em: 14 fev. 2021.

DRAKOULOGKONAS, P.; APOSTOLOU, D. On the selection of process mining tools. **Eletronics** 2021, 10, 451, 2021. Disponível em: <https://doi.org/10.3390/eletronics10040451>. Acesso em: 24 mar. 2021.

EBC. Com o uso de tecnologia, CGU evita prejuízos de R\$ 812 milhões ao Estado, Sistema Alice foi desenvolvido em parceria com o TCU. **Agência Brasil**, publicado em 27/08/2019, por Alex Rodrigues, Brasília, 2019. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2019-08/com-uso-de-tecnologia-cgu-evita-prejuizos-de-r-800-milhoes-ao-estado>. Acesso em: 23 mar. 2021.

ELHAGALY, M. et al. Evolution of compliance checking in process mining discipline. **2019 2nd International Conference on Computing, Mathematics and Engineering Technologies** (iCoMET), Sukkur, Pakistan, p. 1-6, 2019. Disponível em: <https://doi.org/10.1109/ICOMET.2019.8673437>. Acesso em: 16 ago. 2020.

FAZZINGA, B. et al. An ensemble-based approach to the security-oriented classification of low-level log traces. **Expert Systems with Applications**, v.153: 113386, 2020. Disponível em: <https://doi.org/10.1016/j.eswa.2020.113386>. Acesso em: 04 jul. 2020.

FEBRIYANTI, K. D.; SARNO, R.; EFFENDI, Y. A. Fraud detection on event logs using fuzzy association rule learning. *In Proceedings of the 11th International Conference on Information and Communication Technology and System*, ICTS 2017, p. 149-154, 2018. Disponível em: <https://doi.org/10.1109/ICTS.2017.8265661>. Acesso em: 04 jul. 2020.

FOLINO, F.; GUARASCIO, M.; PONTIERI, L. A descriptive clustering approach to the analysis of quantitative business-process deviances. In *Proceedings of the Symposium on Applied Computing (SAC '17)*. Association for Computing Machinery, New York, NY, USA, p. 765–770, 2017. Disponível em: <https://doi.org/10.1145/3019612.3019660>. Acesso em: 04 jul. 2020.

FUJIWARA, R. S.; TSUNODA, D. F. Mineração de processos na identificação de fraudes: um escopo da produção científica. *In Proceedings of the 4th Information Systems in Latin America, ISLA 2020*. Disponível em: <https://aisel.aisnet.org/isla2020/17>. Acesso em: 18 ago. 2020.

GARCIA C. dos S. et al. Process mining techniques and applications – A systematic mapping study. *Expert Systems with Applications*, v. 133, p. 260-295, 2019. Disponível em: <https://doi.org/10.1016/j.eswa.2019.05.003>. Acesso em: 09 ago. 2020.

GENGA, L. et al. APD tool: Mining anomalous patterns from event logs. *In Proceedings of the BPM Demo Track and BPM Dissertation Award*. Sun SITE Central Europe (CEUR workshop proceedings), 2017. Disponível em: http://ceur-ws.org/Vol-1920/BPM_2017_paper_186.pdf. Acesso em: 04 jul. 2020.

HUANG, Z.; CONG, Q.; HU, J. Information system risk auditing model based on process mining. *International Conference on Management Science and Engineering - Annual Conference Proceedings*. p. 39-45, 2012. Disponível em: <https://doi.org/10.1109/ICMSE.2012.6414158>. Acesso em: 04 jul. 2020.

HUDA, S. et al. Identification of process-based fraud patterns in credit application. *In Proceedings of the 2014 2nd International Conference on Information and Communication Technology*, ICoICT 2014, p. 84-89, 2014. Disponível em: <https://doi.org/10.1109/ICoICT.2014.6914045>. Acesso em: 04 jul. 2020.

IRB. **Instituto Rui Barbosa**. Declaração de Moscou 2019 (XXIII INCOSAI), tradução IRB e TCE-PR, 2019. Disponível em: https://irbcontas.org.br/wp-content/uploads/2020/04/Declara%C3%A7%C3%A3o_de_Moscou_2019_-_tradu%C3%A7%C3%A3o_livre.pdf. Acesso em: 23 mar. 2021.

IRB. **Instituto Rui Barbosa**. Matéria “Uso de robôs pelos Tribunais de Contas”, equipe Instituto Rui Barbosa, publicado em 10 de julho de 2020. Disponível em: <https://irbcontas.org.br/uso-de-robos-pelos-tribunais-de-contas/>. Acesso em: 23 mar. 2021.

IYER, N., SAMOCIUK, M. **Fraud and corruption: prevention and detection**. Aldershot: Gower, 2006.

JALALI, H.; BARAANI, A. Genetic-based anomaly detection in logs of process aware systems. *World Academy of Science, Engineering and Technology*, v.64, p. 304-

309, 2010. Disponível em: <https://publications.waset.org/13186/genetic-based-anomaly-detection-in-logs-of-process-aware-systems>. Acesso em: 04 jul. 2020.

JANS, M. et al. A business process mining application for internal transaction fraud mitigation. **Expert Systems with Applications**, v.38, n.10, p. 13351-13359, 2011. Disponível em: <https://doi.org/10.1016/j.eswa.2011.04.159>. Acesso em: 04 jul. 2020.

KEBEDE, M. **Comparative evaluation of process mining tools**. 2015. 56 f. Master's Thesis (Masters of Software Engineering) – University of Tartu, Tartu, 2015.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in software engineering**. Version 2.3. EBSE Technical Report (EBSE-2007-01). Keele University and University of Durham: 2007. Disponível em: https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf. Acesso em 09 nov. 2019.

MARDANI, S.; SHAHRIARI, H. R. A new method for occupational fraud detection in process aware information systems. **10th International ISC Conference on Information Security and Cryptology (ISCISC), 2013**, p. 1-5, 2013. Disponível em: <https://doi.org/10.1109/ISCISC.2013.6767348>. Acesso em: 04 jul. 2020.

MAULIDI, A. When and why (honest) people commit fraudulent behaviours? Extending the fraud triangle as a predictor of fraudulent behaviours. **Journal of Financial Crime**, v. 27 n. 2, p. 541-559, 2020. Disponível em: <https://doi.org/10.1108/JFC-05-2019-0058>. Acesso em: 12 ago. 2020.

MUNOZ-GAMA, L. Conformance checking and diagnosis in process mining: comparing observed and modeled process. **Lecture Notes in Business Information Processing**, v. 270, Cham: Springer, 2016.

NOLLE, T. et al. Analyzing business process anomalies using autoencoders. **Mach Learn** 107, p. 1875–1893, 2018. Disponível em: <https://doi.org/10.1007/s10994-018-5702-8>. Acesso em: 04 jul. 2020.

OKOLI, C. Guia para realizar uma revisão sistemática da literatura. Tradução de David Wesley Amado Duarte; Revisão técnica e introdução de João Mattar. **EaD em Foco**, v. 9:1, 2019. Disponível em: <http://eademfoco.cecierj.edu.br/index.php/Revista/article/view/748>. Acesso em: 09 nov. 2019.

PANE, E. S.; WIBAWA, A. D.; PURNOMO, M. H. Event log-based fraud rating using interval type-2 fuzzy sets in fuzzy AHP. **2016 IEEE Region 10 Conference (TENCON)**, p. 1965-1968, 2016. Disponível em: <https://doi.org/10.1109/TENCON.2016.7848366>. Acesso em: 04 jul. 2020.

PAULRAJ, P.; NEELAMEGAM, A. Improving Business Intelligence Based on Frequent Itemsets Using k-Means Clustering Algorithm. N. Meghanathan et al. (eds.), **Networks and Communications (NetCom2013), Lecture Notes in Electrical Engineering** 284, Chapter 19, 2014. https://doi.org/10.1007/978-3-319-03692-2_19. Acesso em: 04 jul. 2020.

PILON, B. et al. Gaussian process for regression in business intelligence: a fraud detection application. *In Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, v. 3: KMIS, (IC3K 2015), p. 39-49, 2015. Disponível em: <https://www.scitepress.org/Link.aspx?doi=10.5220%2f0005593000390049>. Acesso em: 31 ago. 2020.

PRODANOV, C. C.; FREITAS, E. C. de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2.ed. Novo Hamburgo: Feevale, 2013.

PWC. Tirando a fraude das sombras. Pesquisa global sobre fraudes e crimes econômicos 2018. **PricewaterhouseCoopers Brasil**, 2018. Disponível em: <https://www.pwc.com.br/pt/estudos/assets/2018/gecs-18.pdf>. Acesso em: 20 jun. 2020.

RAHMAWATI, D. et al. Fraud detection on event log of bank financial credit business process using Hidden Markov Model algorithm. **2017 3rd International Conference on Science in Information Technology (ICSITech)**, p. 35-40, 2017. Disponível em: <https://doi.org/10.1109/ICSITech.2017.8257082>. Acesso em: 04 jul. 2020.

RAHMAWATI, D.; SARNO, R. Scalable attack analysis of business process based on decision mining classification. **2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)**, p. 1-6, 2017. Disponível em: <https://doi.org/10.1109/EECSI.2017.8239135>. Acesso em: 04 jul. 2020.

RAHMAWATI, D.; YAQIN, M.; SARNO, R. Fraud detection on event logs of goods and services procurement business process using Heuristics Miner algorithm. **2016 International Conference on Information & Communication Technology and Systems (ICTS)**, p. 249-254, 2016. Disponível em: <https://doi.org/10.1109/ICTS.2016.7910307>. Acesso em: 04 jul. 2020.

RAMOS, C. de A. **Estudo comparativo entre ferramentas de mineração de processos de negócio**. Monografia (Bacharelado em Ciência da Computação), 71 pp., Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019.

RANGA SURI, N. N. R.; NARASIMHA MURTY, M.; ATHITHAN, G. **Outlier detection: techniques and applications (a data mining perspective)**. Intelligent Systems Reference Library, v.155, Cham: Springer Nature, 2019.

SARNO, R. et al. Hybrid association rule learning and process mining for fraud detection. **IAENG International Journal of Computer Science**, v.42, n.2, p. 59-72, 2015. Disponível em: https://www.researchgate.net/publication/280805918_Hybrid_Association_Rule_Learning_and_Process_Mining_for_Fraud_Detection. Acesso em: 04 jul. 2020.

SARNO, R., SINAGA, F., SUNGKONO, K. R. Anomaly detection in business processes using process mining and fuzzy association rule learning. **J Big Data**, v. 7:5, 2020. Disponível em: <https://doi.org/10.1186/s40537-019-0277-1>. Acesso em: 19 ago. 2020.

SILVA, F. Analytical Intelligence in Processes: Data Science for Business. **IEEE Latin America Transactions**, v.16, p. 2240-2247, 2018. Disponível em: <https://doi.org/10.1109/TLA.2018.8528241>. Acesso em: 04 jul. 2020.

TALEB, N. N. **Arriscando a própria pele: assimetria ocultas no cotidiano** (tradução Renato Brett). 1ª ed. Rio de Janeiro: Objetiva, 2018.

TAVARES, G. M. et al. Anomaly detection in business process based on data stream mining. In *Proceedings of the ACM International Conference Proceeding Series*, p. 120-127, 2018. Disponível em: <https://doi.org/10.1145/3229345.3229362>. Acesso em: 04 jul. 2020.

TCEPR. Dados abertos TCE-PR. Consulta de bases, obras municipais. **Tribunal de Contas do Estado do Paraná**, 2021. Disponível em: <https://www1.tce.pr.gov.br/conteudo/dados-abertos-tce-pr-consulta-de-bases/318736/area/54>. Acesso em: 16 fev. 2021.

THIO-AC, A., et al. Blockchain-based system evaluation: the effectiveness of Blockchain on e-procurements. **International Journal of Advanced Trends in Computer Science and Engineering**, v. 8, n. 5, p. 2673-2676, 2019. Disponível em: [doi:10.30534/ijatcse/2019/122852019](https://doi.org/10.30534/ijatcse/2019/122852019). Acesso em: 31 ago. 2020.

VAN DER AALST, W. M. P. et al. Process mining manifesto. In: Daniel F., Barkaoui K., Dustdar S. (eds) Business Process Management Workshops. BPM 2011. **Lecture Notes in Business Information Processing**, v.99. Springer, Berlin, Heidelberg, 2012. Disponível em: https://doi.org/10.1007/978-3-642-28108-2_19. Acesso em: 14 ago. 2020.

VAN DER AALST, W. M. P. How People Really (Like To) Work - Comparative Process Mining to Unravel Human Behavior. HCSE, 2014. Disponível em: https://link.springer.com/content/pdf/10.1007%2F978-3-662-44811-3_25.pdf. Acesso em: 04 jul. 2020.

VAN DER AALST, W. M. P. **Process mining: data science in action**. 2 ed. Berlin: Springer, 2016.

VAN DER AALST, W. M. P.; DE MEDEIROS, A. K. A. Process mining and security: Detecting anomalous process executions and checking process conformance. **Electronic Notes in Theoretical Computer Science**, v.121, SPEC. ISS., p. 3-21, 2005. Disponível em: <https://doi.org/10.1016/j.entcs.2004.10.013>. Acesso em: 20 jul. 2020.

VERSTRAETE, D. **Process mining in practice: comparative study of process mining software**. Thesis (Master of science in the applied economics) – Universiteit Gent, Gent, Belgium, 2014.

VINER, D.; STIERLE, M.; MATZNER, M. A process mining software comparison. **International Conference on Process Mining (ICPM) 2020**. Disponível em: <https://arxiv.org/abs/2007.14038v3>. Acesso em: 24 mar. 2021.

WANG, X.; ZHAO, R.; LI, Y. A fraudulent data simulation method based on generative adversarial networks. **Journal of Physics: Conference Series**, v. 1302, n. 2, 2019.

Disponível em: <https://doi.org/10.1088/1742-6596/1302/2/022089>. Acesso em: 26 ago. 2020.

WARDHANA, I. P. S. P.; DANTES, G. R.; ARYANTO; K. Y. E. Analysis of digital identity transactions with ethereum blockchain in a case study of credit applications in banking. **Journal of Physics: Conference Series**, v. 1516, 2nd International Conference on Vocational Education and Technology (IConVET) 2019, Bali, Indonesia, 2019. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1516/1/012020>. Acesso em: 31 ago. 2020.

XIAO, C.; FREEMAN, D. M.; HWA, T. Detecting clusters of fake accounts in online social networks. *In Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security (AISeC '15)*. Association for Computing Machinery, New York, NY, USA, p. 91–101, 2015. Disponível em: <https://doi.org/10.1145/2808769.2808779>. Acesso em: 13 jun. 2020.

YANG, W.-S.; HWANG, S.-Y. A process-mining framework for the detection of healthcare fraud and abuse. **Expert Systems with Applications**, v.31, n.1, p. 56-68, 2006. Disponível em: <https://doi.org/10.1016/j.eswa.2005.09.003>. Acesso em: 04 jul. 2020.

ZERBINO, P. et al. Process-Mining-enabled audit of Information Systems: methodology and an application. **Expert Systems with Applications**, v.110, 2018. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2018.05.030>. Acesso em: 04 jul. 2020.

APÊNDICE A – DETECÇÃO DE FRAUDES COM APRENDIZADO DE MÁQUINA

Parte integrante do estudo apresentado no Consórcio Mestral do II Seminário Nacional de Gestão da Informação e do Conhecimento, em 03/09/2020, de modo virtual. Na sequência o resumo do artigo e a tabela em que constam os métodos de aprendizado de máquina mais empregados na identificação de fraudes.

DETECÇÃO DE FRAUDES COM APRENDIZADO DE MÁQUINA: ANÁLISE DA PRODUÇÃO CIENTÍFICA INDEXADA PELA SCOPUS

Resumo: Este estudo aborda a Detecção de Fraudes apoiada na utilização do Aprendizado de Máquina, através da análise da evolução das produções científicas. De caráter descritiva e abordagem quantitativa, a pesquisa bibliográfica faz uso de um levantamento no portal de periódicos SCOPUS buscando-se os termos aprendizado de máquina e fraude (e as variações em idiomas), recuperando um total de 748 artigos científicos publicados entre 1992 e 2020. O ápice ocorreu em 2019, com 35,4% (265) das publicações, o crescimento também foi identificado através da análise de recorrência das palavras machine learning e fraud. Os pesquisadores dos Estados Unidos (264) são os que tiveram maior número de publicações, seguidos da Índia (211) e China (170). A tendência da utilização de redes neurais é evidenciada com o aparecimento dos termos deep learning, artificial neural network (ANN), convolutional neural network (CNN) e perceptron no total de 103 das publicações, outros algoritmos como o random forest, árvore de decisão e a máquina de vetores de suporte, aparecem na sequência. Considerações parciais, apontam características do interesse na temática, o grande volume de informações que transitam (exemplo de operação de cartão de crédito e publicação de notícias) relacionados as atividades de fraudes, e necessidade de identificar estas fraudes.

Palavras-chave: Inteligência Artificial. Mineração de Dados. Anomalia. Gestão da Informação. Bibliometria.

A lista de métodos, técnicas ou algoritmos que são utilizados no aprendizado de máquinas é extensa, a busca dos principais termos nos títulos, palavras-chaves e resumos resultou a Tabela 5. Os termos *machine learning* ou *machine learning method* são encontrados em 91,0% das publicações, que identificam de forma generalizada o procedimento utilizado nas pesquisas.

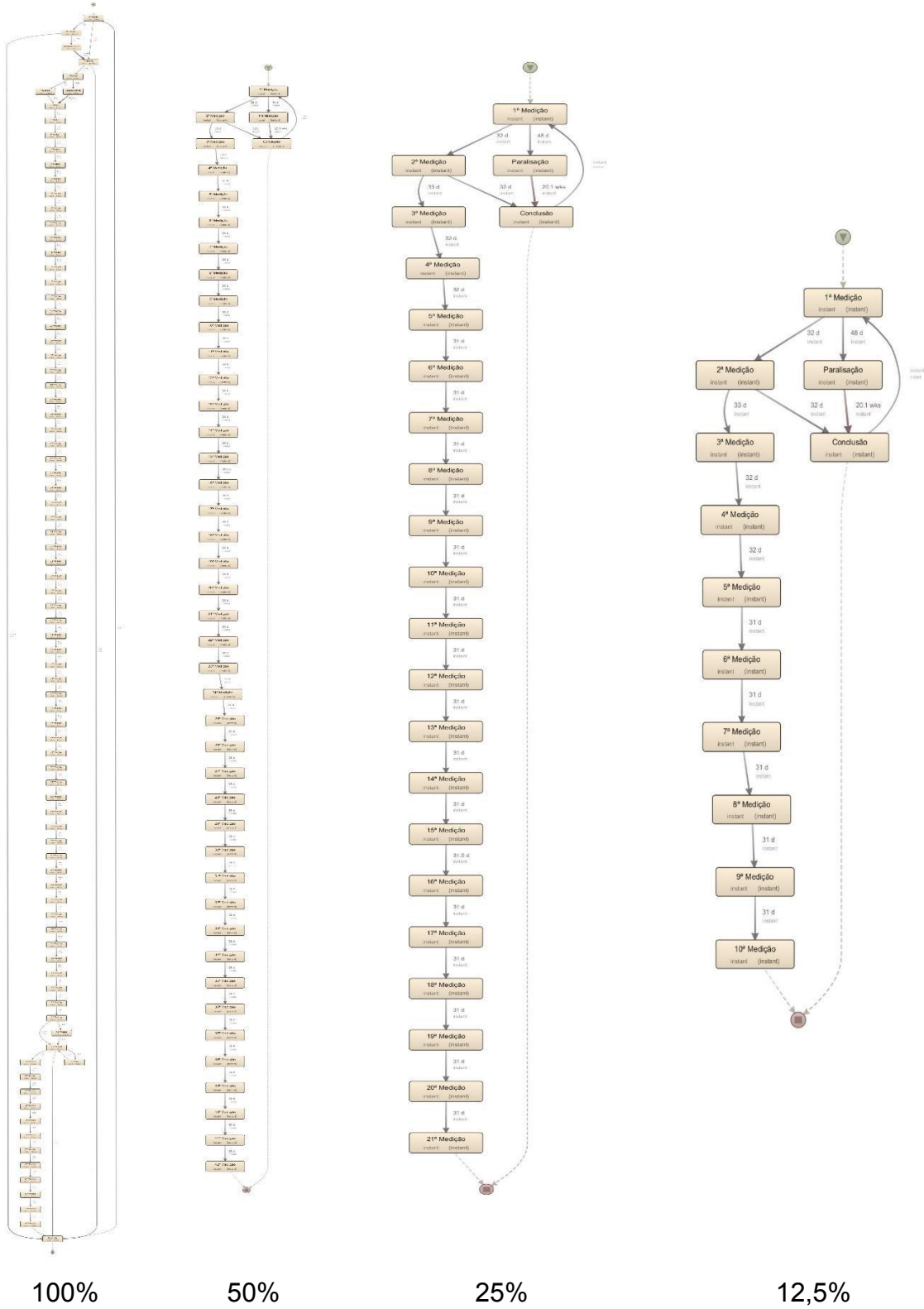
TABELA 5 – ALGORITMOS/TÉCNICAS MAIS EMPREGADOS NAS PUBLICAÇÕES

Técnica / Método / Algoritmo	Número de Publicações	% em relação ao total (748)
Machine Learning or Machine Learning Method	681	91,0%
Neural Network / Deep Learning (Artificial / Convolutional Neural)	103	13,8%
Random Forest	85	11,4%
Decision Tree (C4.5 Algorithm / J48 Algorithm)	75	10,0%
SVM (Support Vector Machine)	73	9,8%
Naive Bayes Algorithm	68	9,1%
Logistic Regression	48	6,4%
K-Nearest Neighbors	22	2,9%
Genetic Algorithm	13	1,7%
Text Mining	9	1,2%
Fuzzy Logic	7	0,9%
Linear Regression	3	0,4%

Fonte: Elaborado pelo autor (2020)

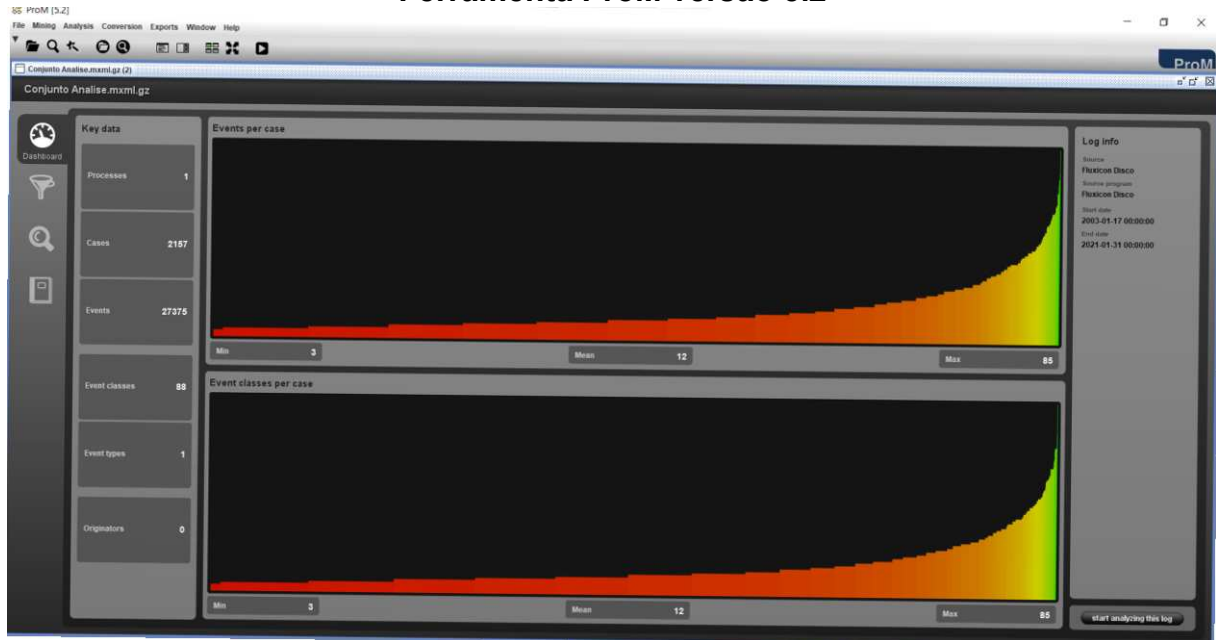
A tendência da utilização de redes neurais (*neural network*) fica evidente na identificação de termos como *deep learning*, *artificial neural network (ANN)*, *convolutional neural network (CNN)* e *Perceptron* abrangendo 103 das publicações, ou 13,8% dos 748 artigos analisados. Em seguida aparecem o algoritmo *Random Forest* (11,4%), os algoritmos de árvore de decisão incluindo o J48 (implementação em Java do método C4.5 na ferramenta Weka) e C4.5 (10,0%) e o algoritmo *Support Vector Machine – SVM* (9,8%).

APÊNDICE B – VISUALIZAÇÃO DE ATIVIDADES NO DISCO (SIMPLIFICAÇÃO DE ATIVIDADES)



APÊNDICE C – TELA DA FERRAMENTA PROM

Ferramenta ProM versão 5.2



Ferramenta ProM versão 6.1



ANEXO I – ELABORAÇÃO DOS DIRECIONAMENTOS FUTUROS DA AUDITORIA (FISCALIZAÇÃO) PÚBLICA DECLARAÇÃO

Parte integrante da Declaração de Moscou, traduzido por: Denise Gomel (TCE-PR) e Nelson Nei Granato Neto (TCE-PR), sendo a última revisão realizada em 22 de outubro de 2019. Relacionam-se os itens 4, 5, 6 e 8 de um total de 10 itens constantes do anexo único da Declaração de Moscou:

RESPONDER DE FORMA EFICAZ ÀS OPORTUNIDADES DECORRENTES DOS AVANÇOS TECNOLÓGICOS

4. As EFS poderiam promover a cultura da disponibilização e abertura dos dados, dos códigos-fonte e dos algoritmos.

Declarações principais:

- Cada vez mais dados gerados pelas administrações públicas são disponibilizados em formatos abertos. Isto cria um ambiente de informação para todos e pode contribuir para uma maior transparência, bem como apoiar a *accountability* governamental.
- As EFS podem exercer um papel importante na informação e orientação dos governos em relação aos benefícios da abertura dos dados públicos. As EFS podem promover a cultura da disponibilização e abertura dos dados produzidos pelo governo, exceto com relação a restrições previstas em leis de confidencialidade ou questões de privacidade.
- As EFS podem promover a abertura dos sistemas informatizados que embasam a tomada de decisões utilizados pelo governo, sob os princípios de códigos-fonte abertos e dados abertos.
- As EFS devem facilitar uma discussão sobre a publicação dos resultados das EFS em formato de dados abertos, quando apropriado.

5. As EFS poderiam objetivar um melhor uso da análise de dados em auditorias (fiscalizações), incluindo estratégias de adaptação como o planejamento destas

auditorias (fiscalizações), o desenvolvimento de equipes experientes em análise de dados e a introdução de novas técnicas na prática de auditoria (fiscalização) pública.

Declarações principais:

- O uso de análise de dados nas EFS é uma inovação necessária, que transforma os dados em uma fonte de recursos para a promoção da eficiência, prestação de contas, eficácia e transparência da administração pública.
- A posição singular das EFS dentro do setor público permite-lhes captar uma grande quantidade de dados das entidades fiscalizadas. O emprego de técnicas de análise de “big data” ao longo do processo de fiscalização permite às E.F.S:
 - Sintetizar dados obtidos de diferentes departamentos, setores, níveis de governo e regiões, o que permite a síntese dos dados obtidos para encontrar soluções para problemas do governo com um todo.
 - Combinar abordagens de coleta de dados (captação própria, externa, ambas) para prover uma atualização regular dos dados e permitir um acompanhamento em tempo real de problemas críticos ou de áreas de maior risco.
- As EFS podem se beneficiar da execução de pesquisas analíticas de “big data” na fase preliminar das auditorias (fiscalizações). Isto encurtará o tempo de trabalho em campo e permitirá um acompanhamento e um monitoramento regulares.
- As EFS podem se beneficiar da replicação de estudos científicos e fortalecer seu trabalho metodológico interno para aplicar correta e apropriadamente os métodos de pesquisa científica. As EFS podem, ainda, se relacionar com instituições acadêmicas para conduzir pesquisas científicas em conjunto.
- Através do estreitamento da cooperação entre as EFS e organizações internacionais relevantes, a INTOSAI pode listar experiências e conhecimento de utilização de “big data” em auditorias (fiscalizações), desenvolver orientações e relatórios de pesquisa relevantes, e encorajar as EFS. em

construir suas próprias capacidades em utilização de “big data” em auditorias (fiscalizações).

REFORÇAR O IMPACTO DAS EFS

6. As EFS podem fomentar uma mentalidade experimental para reforçar a inovação e o desenvolvimento.

Declarações principais:

- Para liderar pelo exemplo, as EFS podem fortalecer suas capacidades inovativas e experimentais, isto é, incluindo fases de aprendizado, teste e avaliação em algumas partes do seu trabalho. Abordagens experimentais nas E.F.S. poderiam acelerar o aprendizado e o desenvolvimento de capacidades ao, sistematicamente, estabelecer e testar hipóteses, e identificar lacunas de conhecimento.
- As EFS poderiam apoiar um uso mais eficiente do conhecimento baseado em experiências na sociedade e uma incorporação mais generalizada da experimentação em programas, estratégias e políticas governamentais.
- A experimentação nas políticas públicas objetiva encontrar maneiras inovadoras para entregar os resultados. Para enfrentar questões complexas e entregar melhores resultados, as EFS podem apoiar abordagens experimentais. Descobrir o que funciona e não funciona nas políticas públicas permite compreender a eficácia das intervenções.

8. As EFS são encorajadas a formar os auditores do futuro capazes de: trabalhar com análise de dados, ferramentas de inteligência artificial e avançados métodos de análise qualitativa; reforçar a capacidade de inovação; atuar como parceiros estratégicos; compartilhar conhecimento e gerar previsões.

Declarações principais:

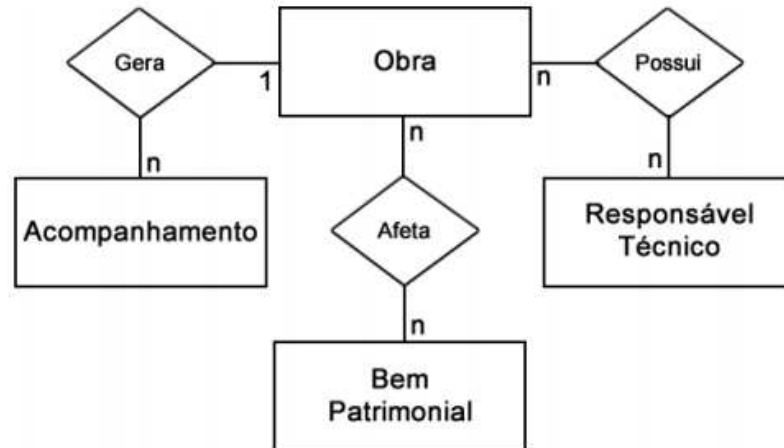
- Garantir a profissionalização do seu pessoal é um objetivo-chave. As competências e as habilidades dos auditores são o maior ativo das EFS.
- As mudanças no ambiente de fiscalização e nas expectativas das partes interessadas moldam os requisitos novos e existentes para o conjunto de habilidades (necessárias) aos auditores das EFS.:
 - *Mentalidade estratégica*: análise de viabilidade, pensamento orientado por hipóteses, identificação de relações causais, orientação por metas, previsão, planejamento estratégico, pensamento sistêmico, priorização;
 - *Habilidades de análise de dados*: trabalho com conjuntos e bases de dados, visualização de dados, apresentação de dados complexos;
 - *Habilidades sociais*: comunicação eficaz, inteligência emocional, construção e manutenção da confiança baseada no profissionalismo, liderança, habilidades para construir consensos.
- Para fortalecer o potencial analítico, as E.F.S. podem estabelecer unidades de análise de problemas específicos (por exemplo, compreensão de risco e gerenciamento de risco, avaliação de programas, entre outros).

ANEXO II – RELACIONAMENTO E DICIONÁRIO DE DADOS DO CONJUNTO OBRAS MUNICIPAIS

TRIBUNAL DE CONTAS DO ESTADO DO PARANÁ

Dados Abertos: “Obras Municipais”

Modelo Entidade Relacionamento (ER)



TCE-PR - DADOS ABERTOS

Dicionário de dados do conjunto: Obras Municipais

Base de dados: **obras_municipais_base_de_dados.csv**

<http://www1.tce.pr.gov.br/conteudo/dados-abertos-tce-pr/318571/area/54>

Nome da coluna	Tipo de dado	Permite Nulo	Descrição
idIntervencao	int	Não	Identificador numérico da intervenção gerado sequencialmente. PK (Primary key) utilizado para os relacionamentos com os outros arquivos de dados do conjunto.
nmPessoa	char(250)	Não	Descrição do ente público responsável pela despesa relativa à intervenção.
nrDocumento	char(15)	Não	Número do Cadastro Nacional de Pessoa Jurídica do ente público.
cdIBGE	char(5)	Não	Código do IBGE do Município de origem do ente responsável pela despesa relativa à intervenção.
nmMunicipio	char(40)	Não	Nome do Município de origem do ente responsável pela despesa relativa à intervenção.
cdIntervencao	smallint	Não	Código da intervenção, que juntamente com o nrDocumento e nrAnoIntervencao formam um conjunto único que identifica a intervenção e equivale a um AK (Alternate Key).
nrAnoIntervencao	numeric(4,0)	Não	Ano de início da execução da intervenção.
dsTipoIntervencao	char(50)	Não	Descrição do tipo da intervenção (execução de obra, projeto ou outras atividades técnicas).
dsClassificacaoIntervencao	char(50)	Não	Descrição da classificação do tipo da intervenção (construção, ampliação, reforma ou outro).
nmIntervencao	char(100)	Não	Denominação que identifica de forma única a intervenção a ser executada e que traduz o tipo de trabalho a ser realizado.
dsTipoObra	char(50)	Não	Descrição do tipo de obra (ex.: edificação, pavimentação, saneamento).
dsClassificacaoObra	char(50)	Não	Descrição da classificação da obra (ex.: creche, hospital, malha viária urbana, dragagem, parque ou praça).
dsObjeto	char(800)	Não	Descrição da intervenção a ser executada.

nrMedida	numeric(11,2)	Não	Quantidade de medida que melhor caracteriza a intervenção.
dsUnidadeMedidaIntervencao	char(50)	Não	Unidade de medida que melhor caracteriza a intervenção (ex.: metro quadrado).
vlIntervencao	numeric(16,2)	Não	Valor total da intervenção. No caso de contratação de toda a execução, corresponde ao valor contratado. No caso de intervenção parcialmente contratada, corresponde ao valor orçado para a execução total da intervenção (contrato + valor da execução direta). No caso de execução direta, corresponde ao valor orçado para a execução total da intervenção (apropriação de todos os insumos, inclusive mão-de-obra).
dtBaseValorIntervencao	smalldatetim e	Não	Data referente ao valor total da intervenção (data do contrato ou data da planilha orçamentária da intervenção).
nrPrazoExecucao	smallint	Não	Prazo inicial previsto para a execução da intervenção, em dias.
dtInicio	smalldatetim e	Não	Data de efetivo início da execução da intervenção.
dsTipoRegimeIntervencao	char(50)	Não	Descrição do regime adotado para a execução da intervenção (direto, indireto ou misto (direto + indireto)).

TCE-PR - DADOS ABERTOS

Dicionário de dados do conjunto: Obras Municipais

Base de dados: **acompanhamentos_base_de_dados.csv**<http://www1.tce.pr.gov.br/conteudo/dados-abertos-tce-pr/318571/area/54>

Nome da coluna	Tipo de dado	Permite Nulo	Descrição
idIntervencao	int	Não	Identificador numérico da intervenção gerado sequencialmente. PK (Primary key) utilizado para os relacionamentos com os outros arquivos de dados do conjunto.
dsOrigemAcompanhamento	char(50)	Não	Origem do acompanhamento (jurisdicionado, TCE-PR ou CREA-PR)
nrAcompanhamento	smallint	Não	Número de controle de acompanhamento (sequencial para cada intervenção).
dtAcompanhamento	smalldatetim e	Não	Data do acompanhamento.
dsTipoAcompanhamento	char(50)	Não	Descrição do tipo de acompanhamento (medição, paralisação, conclusão, cancelamento de intervenção ou cadastro indevido).
nmResponsavel	char(100)	Não	Nome do responsável técnico pelo acompanhamento.
sgTipoDocumentoResponsavel	char(12)	Não	Tipo de documento do responsável técnico pelo acompanhamento.
nrDocumentoResponsavel	char(15)	Não	Número do documento do responsável técnico pelo acompanhamento.
dsObservacao	char(8000)	Sim	Observação do acompanhamento.
dsTipoMedicao	char(50)	Sim	Tipo de medição (de execução indireta - contrato, execução indireta - aditivo ou execução direta).
nrPercentualFisico	numeric(5,2)	Sim	Percentual físico medido (Acumulado).
dsMotivoParalisacao	char(200)	Sim	Motivo informado pelo ente para a paralisação da intervenção.

TCE-PR - DADOS ABERTOS

Dicionário de dados do conjunto: Obras Municipais

Base de dados: **responsaveis_tecnicos_base_de_dados.csv**<http://www1.tce.pr.gov.br/conteudo/dados-abertos-tce-pr/318571/area/54>

Nome da coluna	Tipo de dado	Permite Nulo	Descrição
idIntervencao	int	Não	Identificador numérico da intervenção gerado sequencialmente. PK (Primary key) utilizado para os relacionamentos com os outros arquivos de dados do conjunto.
nmResponsavelTecnico	char(100)	Não	Nome da pessoa com algum tipo de responsabilidade técnica pela intervenção.
sgTipoDocumentoResponsavelTecnico	char(12)	Não	Tipo de documento do responsável técnico (CREA ou CAU).

nrDocumentoResponsavelTecnico	char(15)	Não	Número do documento do responsável técnico.
dsTipoDocumentoOrgaoClassificacao	char(50)	Não	Identifica se a responsabilidade técnica é atribuída através de ART (CREA) ou RRT (CAU).
nrRT	char(20)	Não	Número da ART ou do RRT.
dsTipoResponsabilidadeTecnica	char(50)	Não	Identifica o tipo de responsabilidade técnica (ex.: execução de obra, projeto, consultoria).

TCE-PR - DADOS ABERTOS

Dicionário de dados do conjunto: Obras Municipais

Base de dados: **bens_patrimoniais_base_de_dados.csv**<http://www1.tce.pr.gov.br/conteudo/dados-abertos-tce-pr/318571/area/54>

Nome da coluna	Tipo de dado	Permite Nulo	Descrição
idIntervencao	int	Não	Identificador numérico da intervenção gerado sequencialmente. PK (Primary key) utilizado para os relacionamentos com os outros arquivos de dados do conjunto.
cdBem	char(10)	Não	Código de controle do bem patrimonial.
dsBem	char(250)	Não	Descrição do bem patrimonial que será projetado, criado ou alterado pela intervenção.
dsTipoPropriedadeBem	char(50)	Não	Descrição do tipo de propriedade do bem patrimonial (próprio ou de terceiro).
dsTipoNaturezaBem	char(50)	Não	Descrição do tipo de natureza do bem patrimonial (bem móvel, imóvel ou intangível).
dsTipoCategoriaBem	char(100)	Não	Descrição do tipo de categoria do bem patrimonial (ex.: edificações; terrenos; praças, parques e bosques; ruas, logradouros e estradas).
dsTipoDetalhamentoBem	char(100)	Não	Descrição do tipo de detalhamento do bem patrimonial (ex.: creches, escolas, posto de saúde, ruas, pontes, sistema de esgoto).
dsTipoUtilizacaoBem	char(50)	Não	Identifica se os bens são de uso comum, especial ou dominicais.
dtOperacao	smalldatetime	Não	Representa a data da aquisição ou cadastro do bem.
dtLocalizacao	smalldatetime	Não	Data do registro com a localização do imóvel naquele momento (pois uma rua pode mudar de nome).
nmLogradouro	char(100)	Não	Nome do logradouro.
nrLogradouro	char(9)	Não	Número da localização do bem na rua ou avenida.
dsComplemento	char(20)	Sim	Descrição do complemento.
nmBairro	char(50)	Não	Nome do bairro.
nrCEP	int	Não	Código de Endereçamento Postal - CEP.
nrGrauSul	int	Não	Parte da coordenada geográfica do bem patrimonial, referente à sua localização, em graus, na latitude sul.
nrMinutoSul	int	Não	Parte da coordenada geográfica do bem patrimonial, referente à sua localização, em minutos, na latitude sul.
nrSegundoSul	numeric(4,1)	Não	Parte da coordenada geográfica do bem patrimonial, referente à sua localização, em segundos, na latitude sul.
nrGrauOeste	int	Não	Parte da coordenada geográfica do bem patrimonial, referente à sua localização, em graus, na longitude oeste.
nrMinutoOeste	int	Não	Parte da coordenada geográfica do bem patrimonial, referente à sua localização, em minutos, na longitude oeste.
nrSegundoOeste	numeric(4,1)	Não	Parte da coordenada geográfica do bem patrimonial, referente à sua localização, em segundos, na longitude oeste.
dtMedicao	smalldatetime	Não	Data do levantamento das coordenadas geográficas.