

8-10-2020

## **Mineração de Processos na Identificação de Fraudes: um Escopo da Produção Científica**

Ricardo Shigueru Fujiwara  
*Universidade Federal do Paraná, ricardo.fujiwara@gmail.com*

Denise Fukumi Tsunoda  
*Universidade Federal do Paraná, dtsunoda@gmail.com*

Follow this and additional works at: <https://aisel.aisnet.org/isla2020>

---

### **Recommended Citation**

Fujiwara, Ricardo Shigueru and Tsunoda, Denise Fukumi, "Mineração de Processos na Identificação de Fraudes: um Escopo da Produção Científica" (2020). *ISLA 2020 Proceedings*. 17.  
<https://aisel.aisnet.org/isla2020/17>

This material is brought to you by the Latin America (ISLA) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ISLA 2020 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Mineração de Processos na Identificação de Fraudes: um Escopo da Produção Científica

*Artigo Completo*

**Ricardo Shigueru Fujiwara**  
Mestrando no Programa de Pós-  
Graduação em Gestão da Informação  
Universidade Federal do Paraná  
ricardo.fujiwara@gmail.com

**Denise Fukumi Tsunoda**  
Professora do Departamento de Ciência  
e Gestão da Informação  
Universidade Federal do Paraná  
dtsunoda@gmail.com

## Abstract

This study presents the identification of fraud with support from process mining. The objective is to identify the evolution of scientific production related to the application of process mining in the identification of fraud, through bibliometric analysis in the database Scopus, Web of Science and Ebsco. The research is exploratory, with a quantitative approach, and collection of information from bibliographic research, with subsequent bibliometric analysis. The search resulted in 35 articles published in periodicals and conference proceedings, distributed in 28 sources, being the first publication in 2005. The years 2017 and 2018 obtained the highest number of publications. Researchers from Indonesia presented the highest scientific contribution in the period, followed by Italy and Holland, with Brazil appears in the 5th position. The keywords that demonstrated the main methods applied in mining: genetic algorithms, clustering rules and association rules. The use of the open source tool ProM is identified in several articles, being used in process mining.

## Keywords

Process mining, fraud, process discovery, event logs, bibliometrics.

## Resumo

Este estudo apresenta a identificação de fraudes com suporte da mineração de processos. O objetivo é identificar a evolução da produção científica relacionada à aplicação da mineração de processos na identificação de fraudes, através de análises bibliométricas nas bases de dados Scopus, Web of Science e Ebsco. A pesquisa é exploratória, com abordagem quantitativa e coleta de informações de pesquisa bibliográfica, com posterior análise bibliométrica. A busca resultou em 35 artigos publicados em periódicos e anais de congressos, distribuídos em 28 fontes, sendo a primeira publicação em 2005. Os anos de 2017 e 2018 obtiveram o maior número de publicações. Pesquisadores da Indonésia apresentaram a maior contribuição científica no período, seguidos pela Itália e Holanda, com o Brasil na 5ª posição. As palavras-chave que demonstraram os principais métodos aplicados na mineração: algoritmos genéticos, regras de agrupamento e regras de associação. O uso da ferramenta de código aberto ProM é identificado em vários artigos, sendo usado na mineração de processos.

## Palavras-chave

Mineração de processos, fraude, descoberta de processo, registro de eventos, bibliometria.

## Introdução

O volume de registros (*logs*) de processos, operações ou transações que trafegam dentro e fora das organizações é porta para ações fraudulentas. Em pesquisa global sobre fraude bancária realizada pela empresa de auditoria KPMG, 70% dos entrevistados indicaram usar sistemas de monitoramento de riscos em tempo real, 63% usam uma combinação de regramentos e algoritmos de aprendizado de máquina na detecção de fraudes, 67% utilizam-se da biometria física e 33% da comportamental (KPMG 2019).

Em outra pesquisa realizada pela empresa PwC, considerando um período de dois anos, aponta que 38% dos entrevistados tiveram perda monetária causada pela fraude de até US\$ 100.000, para 37% as perdas foram entre US\$ 100.000 e US\$ 5 milhões e para 14% as perdas ultrapassaram os US\$ 5 milhões. O mesmo estudo aponta o crescimento de empresas impactadas com a fraude, no Brasil o percentual que era de 12% (2016) na pesquisa anterior passando para 50% (2018), o índice global passou de 36% (2016) para 49% (2018), provocando o aumento das despesas no combate à fraude (PWC 2018).

Aplica-se vários métodos de identificação de fraudes, de forma combinada, desde processos manuais baseadas em denúncias de clientes, em processos de auditoria, chegando a processos informatizados como monitoramento em tempo real e técnicas de aprendizado de máquina (KPMG, 2019, p.14). Os crimes cibernéticos causam impactos econômicos nas pessoas e empresas, através do comprometimento dos sistemas de segurança; como consequência surge a necessidade de proteger as redes de sistema, através da identificação de vulnerabilidades e aprimoramento dos modelos de segurança (Hong and Kim 2016). A segurança dos sistemas de informações das organizações está exposta a violações que podem ser conduzidas por invasores (*hackers*), por crimes de colarinho branco, envolvendo temas como governança corporativa, segurança de computadores, procedimentos de auditoria, onde é possível a identificação ausência ou presença de certos comportamentos (Van der Aalst and de Medeiros 2005)

Os processos das organizações evoluem com o tempo, novos processos são criados, provocando a necessidade de descobrir e entender as modificações ocorridas, informações subsidiam futuras readequações nos modelos e sistemas (Cook and Wolf 1998). A mineração de processos oferece a capacidade de extrair modelos a partir de um conjunto de registros de transacionais (*transactional logs*), de forma não tendenciosa permitindo encontrar falhas nos processos investigados, identificar fraudes, com isso auxiliando no monitoramento de controles internos (Jan et al. 2011). Além de considerar que a geração de dados é dinâmica e contínua, necessitando de análises em tempo real ou durante a execução para identificação de casos anômalos, não somente registros de eventos completos (Tavares et al. 2018).

A transparência sobre os mecanismos empregados em filtros de *spam*, detecção de fraudes com cartão de crédito, análise de seguros e empréstimos, dependem de algoritmos computacionais, que podem ter seu acesso dificultado por alguns motivos: o sigilo corporativo, que visa manter a proteção dos segredos comerciais; a desconhecimento técnico, que considera a necessidade habilidade específicas para interpretação dos algoritmos computacionais; e por fim a complexidade ou escala de aplicação de plataformas formados por multicomponentes, que proporciona o desconhecimento da operacionalidade do sistema como um todo (Burrell 2016).

O crescimento dos sistemas organizacionais, juntamente com a quantidade de dados e de transações que operam em meios digitais disponibilizam uma oportunidade para os fraudadores, e de outro lado o conhecimento técnico para identificação das fraudes, muitas vezes é restrito a certos grupos corporativos, não sendo de fácil acesso a pequenas e médias organizações. A busca de produções científicas sobre o assunto “mineração de processo” e “fraude” é uma fonte que visa identificar uma oportunidade de campo de estudo para pesquisadores, deste modo, objetiva-se identificar a evolução da produção científica relacionada à aplicação da mineração de processos na identificação de fraudes, por meio de uma análise bibliométrica nas base de dados Scopus, Web of Science e Ebsco.

O estudo inicia-se pela introdução, na sequência apresentam-se o conceito de fraude e mineração de processos, passando para o percurso metodológico, apresentação dos resultados obtidos e respectiva análise destes. Finaliza-se com a conclusão, agradecimentos e referências.

## Fraude

Os atos ilegais que visam enganar, encobrir, independente de violência ou ameaça física conceituam a fraude, sendo um fenômeno que provoca grandes prejuízos as organizações e em alguns momentos são de difícil detecção, devido a participação de funcionários e de conhecedores do modo de operação dos processos. O crescimento do volume de geração e armazenamento de dados, tornam mais difícil a detecção de fraudes; como consequência surge a necessidade da aplicação de sistemas informatizados para a análise de grandes conjuntos de dados e auxiliar de investigação dos auditores na detecção de fraudes e anomalias (Bernardino, Pedrosa and Laureano 2018).

Desvios ou anomalias em processos de negócios podem ter origem em erros ou variações nas operações, como também podem ser ocasionadas por comportamentos fraudulento, que segundo a teoria do triângulo da fraude pode ser originada por três características: motivação ou pressão, oportunidade (conhecimento) e racionalização. As fraudes em processos de negócios podem ser identificadas previamente pelos controles internos, através de análises nas atividades ignoradas ou com sequências erradas no processo, tempo de execução fora do padrão ou a execução de atividade por pessoa não responsável, são alguns dos exemplos que indicam anomalias. A mineração de processos e a mineração de dados fornecem soluções detecção de anomalias, e consequentemente de fraudes. (Sarno, Sinaga and Sungkono 2020)

## Mineração de Processos

Os sistemas de informações que fornecem suporte aos processos de negócios geram muitos registros que refletem cada ação ou atividade executada dentro da empresa, porém estes dados nem sempre são usados pelas organizações como análise dos seus processos, nesta linha, a mineração de processos objetiva diagnosticar os processos através da mineração dos registros de eventos, transformando dados em conhecimento (Jan et al. 2011).

A descoberta do processo (modelo do processo) é uma etapa importante na caracterização do processo, não somente para identificar o percurso ou etapas contidas nele, mas também para identificar alterações ou caminhos anormais. A metodologia do diagnóstico do processo segue algumas etapas: a preparação dos registros (*logs*), identificação ou inspeção dos registros (qual a informação contida nos registros), análise dos fluxos de controle (identificação dos processos), análise de desempenho (tempo, desvios, gargalos) e análise de função, onde procura-se a identificação do responsável pela execução da atividade (Jan et al. 2011).

A análise de conformidade compara o comportamento observado nos registros de eventos com o modelo de processo, podendo ser avaliados quanto as dimensões de adequação, precisão, generalização e simplicidade. As ineficiências ou possíveis fraudes podem ser identificadas ao verificar o atendimento a conformidade, além de fornecer informações para a auditoria ou permitir que a readequação dos modelos de processos (Adriansyah 2015). A fraude em processos pode ser identificada a partir de atributos de identificação como: pular etapas, execução de atividades diferente do modelo do processo, com tempo diferenciado ou executado por pessoa não responsável pela atividade (Huda et al. 2014)

## Percurso Metodológico

A pesquisa é exploratória, com o intuito de proporcionar maiores informações sobre o assunto de investigação; possui abordagem quantitativa, visando a quantificação das informações, com a aplicação do método de pesquisa bibliográfica, com ênfase artigos científicos (Prodanov and Freitas 2013, pp. 52, 54 e 69).

A Tabela 1 apresenta os resultados das pesquisas realizadas em 03 de julho de 2020, nas bases Scopus, Web of Science e Ebsco. A busca na Ebsco foi realizada na Computer & Applied Sciences Complete, Information Science & Technology Abstracts (ISTA) e na Library, Information Science & Technology Abstracts (LISTA).

Estratégia de pesquisa	“mineração de processos” ou “process mining”	(“mineração de processos” ou “process mining”) e (“fraude” ou “fraud”)
Scopus	3.407	35
Web of Science	1.863	27
Ebsco	488	8
<b>Soma</b>	<b>5.758</b>	<b>70</b>

**Tabela 1. Resultados da conforme termos**

Fonte: Elaborado pelos autores (2020)

A estratégia de pesquisa foi a busca pelos termos em português “mineração de processos” e “fraude” e os respectivos termos em inglês “process mining” e “fraud” nos títulos, palavras-chave e resumos. Não se utilizou um limitador de ano de publicação, com intuito de identificar todas as produções e a sua evolução até a data da pesquisa, quanto ao tipo de documento contemplou-se artigos publicados em periódicos e em anais de conferência, e escritos em português, espanhol ou inglês. Os parâmetros de consulta com os códigos de campos e operadores, resultou em:

((“mineração de processos” OR “process mining”) AND (“fraud\*))

Dos 70 documentos resultantes da soma das pesquisas individuais, 34 eram duplicados e um sem indicação de autoria (NA NA), restando 35 para a análise. Na base Ebsco os 8 artigos, já constavam na pesquisa realizada na Scopus, e dos 27 resultados encontrados na Web of Science, somente dois não constavam nos resultados da Scopus.

A análise métrica foi realizada com a utilização do Biblioshiny (interface gráfica do pacote Bibliometrix, produzido em linguagem R) e do editor de planilhas Microsoft Excel. Utilizou-se a exportação dos resultados no formato BibTex para a importação no Biblioshiny, e deste exportou-se para o Excel para realização de outras análises.

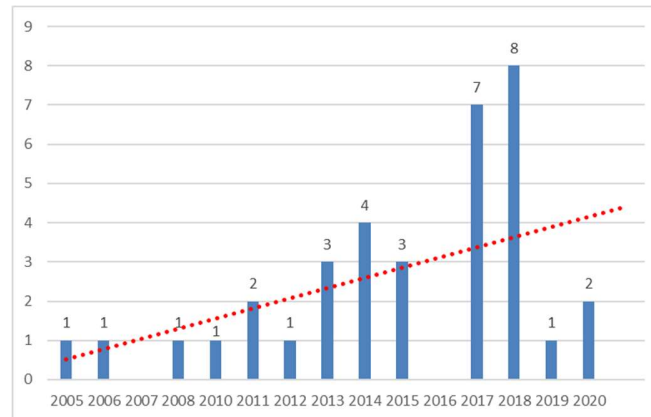
Na identificação dos campos-colunas no arquivo exportado do Biblioshiny, utilizou-se como referencial o artigo “bibliometrix: An R-tool for comprehensive science mapping analysis” (ARIA, CUCCURULLO, 2017, p. 959), que relaciona a codificação e o respectivo conteúdo:

- AU – Authors (Autores)
- TI - Document Title (Título do documento)
- DE - Authors’ Keywords (Palavras-chaves do autor)
- ID - Keywords associated by Scopus database (Palavras-chaves relacionadas pela base de dados Scopus)
- AB – Abstract (Resumo)

## Apresentação e Discussão dos Resultados

A análise compreende 35 artigos, sendo 15 artigos de periódicos e 20 publicados em anais de conferências. No Gráfico 1, temos a evolução das publicações sobre o tema mineração de processos relacionado a identificação de fraudes, compreendendo o período de 2005 a 03 de junho de 2020. O volume de publicações durante o período de 15 anos não é expressivo, na média anual de 2,33, pode caracterizar oportunidade de novas pesquisas.

Temos dois anos concentrando 43% do total, que são 2017 (7) e 2018 (8); em vários anos temos uma única publicação (2005, 2006, 2008, 2010, 2012, 2019), também se nota a ausência de publicações nos anos de 2007 e 2016. Os 7 artigos do ano de 2017 foram resultados de conferências, e em 2018 foram 4 publicações em anais de conferências, do total de 8, que é um indicativo da importância dos eventos científicos para o estudo na temática mineração de processos e fraude.

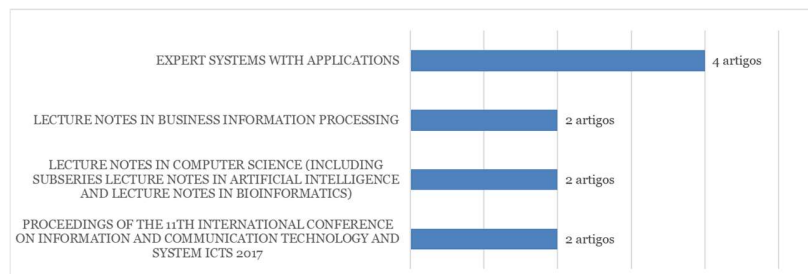


**Gráfico 1. Evolução da produção científica**

Fonte: Elaborado pelos autores (2020)

Os 35 documentos estão publicadas em 28 fontes de pesquisas. A *Expert System with Applications* é que apresenta o maior número de artigos publicados, no total de 4, nos anos de 2006, 2011, 2018 e 2020 e, na sequência, aparecem três fontes de pesquisa com duas publicações cada, sendo que as publicações da *Lecture Notes In Business Information Processing* são referentes a anais de conferências. A outra publicação do ano de 2020 tem como fonte o *Journal of Big Data*, e a primeira publicação (2005) na *Electronic Notes in Theoretical Computer Science*.

Realizada em Surabaya na Indonésia a *11th International Conference On Information And Communication Technology And System (ICTS 2017)*, foi organizada pelo Departamento de Informática do *Institut Teknologi Sepuluh Nopember*, sendo publicadas no IEEE Xplore e indexadas na Scopus. O Gráfico 2 apresenta as fontes com maiores números de publicações, sendo que as 24 fontes restantes apresentam uma única publicação cada.



**Gráfico 2. Fontes com maiores números de publicações**

Fonte: Elaborado pelos autores (2020)

Os 6 pesquisadores mais produtivos estão relacionados na Tabela 2, e os 7 mais citados na Tabela 3. O professor Riyanarto Sarno, do *Institut Teknologi Sepuluh Nopember*, na Indonésia, aparece como mais produtivo (8) e o segundo mais citado (48); o seu artigo mais referenciado *Identification of Process-based Fraud Patterns in Credit Application*, em coautoria com Solichul Huda (Universitas Dian Nuswantoro), Tohari Ahmad (Institut Teknologi Sepuluh Nopember) e Heru Agus Santoso (Universias Dian Nuswantoro), aborda a identificação de fraudes em solicitações de empréstimos, que podem ocorrer na falsificação de documentos ou nos próprios procedimentos, e como resultado da pesquisa propõem 10 atributos que contribuem para utilização de um método de detecção de fraude em processos (Huda et al. 2014).

O primeiro autor mais citado (157) é o professor Willibrordus M. P. Van der Aalst, da RWTH Aachen University, na Alemanha, que aparece em segundo na lista de mais produtivos (5). Seu artigo *Process Mining and Security: Detecting Anomalous Process Executions and Checking Process Conformance*, de coautoria com a pesquisadora Ana Karla Alves de Medeiros, é o artigo mais antigo recuperado na consulta

por termos (2005), abordando o uso de técnicas de mineração de processos para a identificação de padrões ou ausência de padrões nas inspeções em trilhas de auditorias (Van Der Aalst and De Medeiros 2005).

Autores	Artigos
Sarno, R.	8
Van Der Aalst, W. M. P.	5
Folino, F.	4
Pontieri, L.	4
Rahmawati, D.	3
Guarascio, M.	3

**Tabela 2. Autores mais produtivos**

Fonte: Elaborado pelos autores (2020)

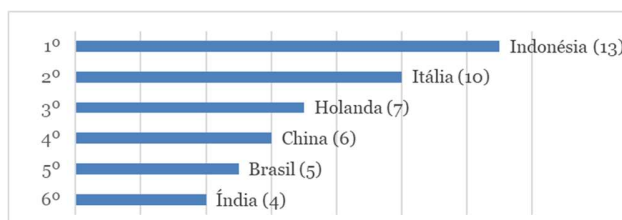
Autores	Citações
Van Der Aalst, W. M. P.	157
Sarno, R.	48
Weijters, A. J. M. M.	33
Wainer, J.	30
Bezerra, F.	26
De Medeiros, A. K. A.	25
Jans, M.	25

**Tabela 3. Autores mais citados**

Fonte: Elaborado pelos autores (2020)

Os 6 países com maiores contribuições científicas no mundo estão identificados no Gráfico 3. Os resultados apontam que os pesquisadores da Indonésia (13) foram os que mais contribuíram nos estudos que conectam a mineração de processos e a identificação de fraudes, em sequência temos a Itália (10), Holanda (7), o Brasil (5) aparece na 5ª posição. Os pesquisadores brasileiros que estão nesta listagem pertencem a Universidade Estadual de Campinas (UNICAMP), Universidade Federal de São Carlos (UFSCar) e Universidade Estadual de Londrina (UEL), e abordam algoritmos para detecção de anomalias em processos (Bezerra and Wainer 2011), realizam um mapeamento sistemático das pesquisas envolvendo big data e inteligência analítica, identificando pesquisas na área de auditoria de fraudes, descoberta e verificação de conformidades (Silva 2018) e aplicam técnicas de agrupamento (*clustering*) em fluxos de eventos, para descoberta de eventos anômalos (Tavares et al. 2018).

Em termos de países com maior número de citações, a Holanda aparece na primeira posição com 164 citações, seguido da China (135), Bélgica (77) e na quarta posição o Brasil com 39 citações.



**Gráfico 3. Contribuição Científica por País ou Região**

Fonte: Elaborado pelos autores (2020)

Os cinco artigos mais referenciados estão listados na Tabela 4. “*A process-mining framework for the detection of healthcare fraud and abuse*” tem como autores Wan-Shiou Yang (National Changhua University of Education) e San-Yih Hwang (National Sun Yat-sen University), é o artigo mais referenciado, abordando a identificação de desvios ou fraudes nos procedimentos clínicos com a utilização da mineração de processos, com objetivo de substituir os processos de análise manual por sistemas automáticos, a pesquisa utilizou-se de dados reais do Programa Nacional de Seguro de Saúde, de Taiwan (Yang and Hwang 2006).

A pesquisa de Jans, Van der Werf, Lybaert e Vanhoof (2011) realiza a aplicação da mineração de processos em um estudo de caso em uma instituição financeira, analisando as aquisições (ciclo de compras), que envolve desde a criação do pedido, liberação, entrada do produto, faturamento chegando ao pagamento. A aplicação da mineração de processos na identificação de fraudes, nas transações internas, obteve resultados positivos quanto a contribuição desta técnica nas práticas comerciais (Jan et al. 2011). “*Measuring precision of modeled behavior*”, neste estudo os pesquisadores propõem um método para medir a precisão dos modelos de processos de forma a comparar com os registros de eventos produzidos nos mesmos processos (Adriansyah et al. 2014)

Documento (constante na amostra de 35 artigos)	Número de Citações
Yang, W.-S., and Hwang, S.-Y. 2006. "A process-mining framework for the detection of healthcare fraud and abuse".	131
Van Der Aalst, W. M. P., and De Medeiros, A. K. A. 2005. "Process mining and security: Detecting anomalous process executions and checking process conformance".	84
Jans, M., Van Der Werf, J. M., Lybaert, N., and Vanhoof, K. 2011. "A business process mining application for internal transaction fraud mitigation".	77
Adriansyah, A., Munoz-Gama, J., Carmona, J., van Dongen, B. F., and Van Der Aalst, W. M. P. 2014. "Measuring precision of modeled behavior".	54
Sarno, R., Dewandono, R. D., Ahmad, T., Naufal, M. F., and Sinaga, F. 2015. "Hybrid association rule learning and process mining for fraud detection".	43

**Tabela 4. Os cinco artigos mais referenciados no período**

Fonte: Elaborado pelos autores (2020)

Os quatro artigos mais citados nas publicações estão listados na Tabela 5. O primeiro deles, "*Discovering models of software processes from event-based data*", de Cook e Wolf (1998) que analisa a aplicação de três métodos para análise de dados de processos (Ktail, Markov e RNet) no intuito de fornecer suporte para a descoberta de processos e o estudo de Cortes e Vapnik (1995) que aborda a aplicação de rede de vetores de suporte (*support-vector network*) na resolução de problemas de classificação, transformando vetores não-lineares em uma superfície linear de decisão.

O artigo "*Workflow Mining: Discovering Process Models From Event Logs*" aborda a descoberta de modelos de fluxos de trabalhos através da análise dos registros do fluxo existentes, as técnicas de mineração de processos foram em duas aplicações reais, a primeira nos procedimentos de pacientes com doença vascular periférica e em processamento de multas de uma agência de cobrança, neste último, o processo de mineração identificou 99 tarefa em 130.136 casos, o resultado teve a validação da agência holandesa *CJIB - Centraal Justitieel Incassobureau*, fonte dos registros (Van Der Aalst, Weijters and Maruster 2004).

Documento citado nas referências (artigos)	Número de Citações
Cook, J. E., and Wolf, A. L. 1998. "Discovering models of software processes from event-based data".	4
Van Der Aalst, W.M.P., Weijters, T., Maruster, L. 2004. "Workflow mining: discovering process models from event logs".	4
Cortes, C., and Vapnik, V. 1995. "Support-vector networks".	3
Van Der Aalst, W. M. P., and De Medeiros, A. K. A. 2005. "Process mining and security: Detecting anomalous process executions and checking process conformance".	3

**Tabela 5. Os quatro artigos mais referenciados nos artigos recuperados**

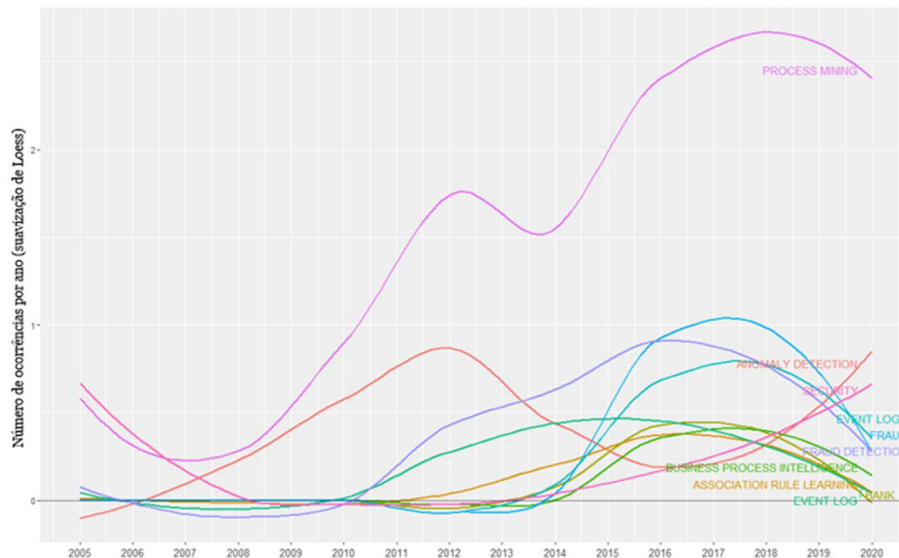
Fonte: Elaborado pelos autores (2020)

Das escolhas dos autores de palavras-chaves, a de maior ocorrência é *process mining* citada 24 vezes, seguida de *anomaly detection* (6), *fraud detection* (6), *fraud* (5), *event logs* (4). A Figura 1 apresenta a evolução da utilização das 10 palavras-chave (termos) com maior ocorrência.

O método de aprendizagem de regras de associação é utilizado em mineração de dados, no estudo de Sarno, Dewandono, Ahmad, Naufal e Sinaga (2015) os pesquisadores propõem um método híbrido de mineração de processos e mineração de dados, o primeiro identificaria as variáveis de fraude gerando regras de associação para identificação da fraude. Outros dois artigos também abordam a utilização integrada de



mineração de processos com mineração de dados, no primeiro aplicou-se aprendizagem de regras de associação onde o método produziu melhores resultados onde a fraude possuía alto de nível de confiança e a aprendizagem de regras de associação difusa (*fuzzy*) onde o nível de confiança na fraude é baixo, além de proporcionar redução dos falsos positivos (Sarno, Sinaga and Sungkono 2020). No segundo artigo aplicou-se o método mineração de processos para identificar anomalias nos processos de negócios, e o método de regra de associação difusa (*fuzzy*) para detectar as fraudes (Febriyanti, Sarno and Effendi 2018). Nos três artigos utilizou-se na mineração de processos o *framework* de código aberto ProM (<http://www.processmining.org/> ou <http://promtools.org/>) em linguagem Java.



**Figura 1. Evolução da utilização das 10 palavras-chaves com maior ocorrência**

Fonte: Elaborado pelos autores (2020)

A Figura 2 apresenta o relacionamento das principais palavras-chaves indicadas pelo autor. Identificamos os relacionamentos entre a mineração de processos com a finalidade de detecção de anomalias, detecção de fraudes e a segurança, objeto desta pesquisa. Além a aplicação em sistemas de conhecimento de processo (*process aware systems*) e os registros dos eventos (*event logs*) fonte de análise da mineração de processo. Aparecem entre as palavras algumas das técnicas aplicadas como a aprendizagem de regras de associação (*association rule learning*), agrupamento (*clustering*), algoritmos genéticos (*genetic algorithm*) e mineração heurística (*heuristic miner*).



**Figura 2. Relacionamento das 13 principais palavras-chaves**

Fonte: Elaborado pelos autores (2020)

A aplicação do algoritmo de agrupamento foi identificada no artigo “Anomaly detection in business process based on data stream mining”, que utilizou o algoritmo *DenStream* incluso na ferramenta *Massive Online Analysis* (<https://moa.cms.waikato.ac.nz/>) desenvolvido em código aberto na linguagem Java, fazendo parte do projeto Weka da Universidade de Waikato, na Nova Zelândia. Jalali and Baraani (2010) abordam

o algoritmo genético aplicado na mineração de processos, esta solução visa descobrir um modelo de processo em registros de eventos e posteriormente detectar traços anômalos com base no modelo apresentado, no estudo os autores também utilizaram o ProM.

No estudo de Caesarita, Sarno e Sungkono (2018), aplicam de forma comparativa os algoritmos de mineração de processos alpha ++ e o minerador heurístico, ambos disponíveis no *framework* ProM que resultou no melhor desempenho do algoritmo heurístico, devido a sua característica de cálculo de frequência que possibilita a identificação de ruídos nos dados.

## Conclusão

Este estudo apresentou a evolução das produções científicas relacionadas as técnicas de mineração de processos (*process mining*) na identificação de fraudes. A consulta em três bases de dados retornou 35 artigos (2005-2020), com ápice nos anos de 2017 e 2018, e nenhuma publicação nos anos de 2007 e 2016. As informações obtidas e analisadas, até o momento, permitiram identificar poucas contribuições científicas no período, apontar quais as publicações mais citadas e mais referenciadas, identificar as principais fontes de publicação, os países com maiores participação científica e as palavras mais utilizadas.

Nesta abordagem preliminar com intuito de capturar o panorama da produção científica, vislumbra-se uma oportunidade e a necessidade do aprofundamento dos estudos nesta temática, a leitura de alguns artigos demonstrou que a necessidade surge com o aumento do volume de dados, a sua forma dinâmica e escalável. Esta necessidade também pode ser exemplificada na dificuldade do tratamento de dados relacionados as transações com cartão de crédito, nos diversos processos internos das organizações, pedidos de indenização ou ressarcimento em seguradoras e operadoras de saúde, são algumas das possibilidades elencadas pelos pesquisadores para a utilização da mineração de processos.

A análise compreendeu a identificação dos principais artigos com a sua leitura, onde encontra-se estudos comparativos, buscando analisar a eficiência dos modelos disponíveis, outros utilizam-se de comparativos de estudos publicados, e métodos multicamadas ou híbridos com a utilização de mineração de dados para gerar regras de associação e classificação. Nota-se nos artigos a tendência de utilização de ferramentas de código aberto como o ProM, específico para utilização na mineração de processos. A disponibilização de *framework* gratuito é um facilitador nas pesquisas em universidades públicas aqui no Brasil.

A oportunidade de pesquisar o assunto surge com a identificação de que existe uma concentração dos estudos em alguns autores e universidades, além da participação em termos globais de poucos países. A iniciativa de eventos direcionados promoveu um aumento no número de publicações, o que é indicativo de facilitador no desenvolvimento de pesquisas nesta temática. A relevância também pode ser observada diariamente nos noticiários, como em fraudes de licitações, desvios de dinheiro, sonegação, corrupção, operações policiais, *fake news*, e mais uma infinidade de situações. A aplicação em processos de auditorias, ou como controle e monitoramento em um sistema de governança, gestão de riscos e integridade (*compliance*) são campos que podem ser explorados com a mineração de processos, e a identificação de anomalias e fraudes.

Um limitador do estudo foram os poucos artigos recuperados na consulta, assim em futuros estudos a estratégia de pesquisa pode abranger outros termos como a mineração de dados, mineração de textos, aprendizado de máquina, ou inteligência artificial vinculando-se a análise de registros de eventos e aos modelos de processos.

## REFERÊNCIAS

- Adriansyah, A., Munoz-Gama, J., Carmona, J., van Dongen, B. F., and Van Der Aalst, W. M. P. 2014. "Measuring precision of modeled behavior," *Information Systems and e-Business Management* (13:1), pp. 37-67. (doi:10.1007/s10257-014-0234-7)
- Bernardino, D., Pedrosa, I., and Laureano, R. M. S. 2018. "Analytical methods for auditing and anomaly/fraud detection," in *Proceedings of the Iberian Conference on Information Systems and Technologies, CISTI 2018*. (doi:10.23919/CISTI.2018.8399429)

- Bezerra, F., and Wainer, J. 2011. "Fraud detection in process aware systems," *International Journal of Business Process Integration and Management* (5:2), pp. 121-129. (doi:10.1504/IJBPIIM.2011.040204)
- Burrell, J. 2016. "How the machine 'thinks': Understanding opacity in machine learning algorithms," *Big Data and Society* (3:1). (doi: 10.1177/2053951715622512)
- Caesarita, Y., Sarno, R., and Sungkono, K. R. 2018. "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*, pp. 143-148. (doi:10.1109/ICTS.2017.8265660)
- Cook, J. E., and Wolf, A. L. 1998. "Discovering models of software processes from event-based data," *ACM Transactions on Software Engineering and Methodology* (7:3), pp. 215-249. (doi:10.1145/287000.287001)
- Cortes, C., and Vapnik, V. 1995. "Support-vector networks," *Machine Learning* (20:3), pp. 273-297. (doi:10.1023/A:1022627411411)
- Febriyanti, K. D., Sarno, R., and Effendi, Y. A. 2018. "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*, pp. 149-154. (doi:10.1109/ICTS.2017.8265661)
- Hong, J. B., and Kim, D. S. 2016. "Towards scalable security analysis using multi-layered security models," *Journal of Network and Computer Applications* (75), pp. 156-168. (doi:10.1016/j.jnca.2016.08.024)
- Huda, S., Ahmad, T., Sarno, R., and Santoso, H. A. 2014. "Identification of process-based fraud patterns in credit application," in *Proceedings of the 2014 2nd International Conference on Information and Communication Technology, ICoICT 2014*, pp. 84-89. (doi:10.1109/ICoICT.2014.6914045)
- Jalali, H., and Baraani, A. 2010. "Genetic-based anomaly detection in logs of process aware systems," *World Academy of Science, Engineering and Technology* (64), pp. 304-309.
- Jans, M., Van Der Werf, J. M., Lybaert, N., and Vanhoof, K. 2011. "A business process mining application for internal transaction fraud mitigation," *Expert Systems with Applications* (38:10), pp. 13351-13359. (doi:10.1016/j.eswa.2011.04.159)
- KPMG. 2019. "Pesquisa global sobre fraude bancária - a ameaça multifacetada da fraude: os bancos estão prontos para enfrentar este desafio?," *KPMG Internacional*, maio/2019. ([https://assets.kpmg/content/dam/kpmg/br/pdf/2019/08/br-pesquisa\\_global\\_de\\_fraude.pdf](https://assets.kpmg/content/dam/kpmg/br/pdf/2019/08/br-pesquisa_global_de_fraude.pdf))
- Prodanov, C. C., and Freitas, E. C. de. 2013. *Metodologia do Trabalho Científico [recurso eletrônico]: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico*, 2. ed., Novo Hamburgo: Feevale. ([www.feevale.br/editora](http://www.feevale.br/editora)).
- PWC. 2018. "Tirando a fraude das sombras. Pesquisa global sobre fraudes e crimes econômicos 2018," *PricewaterhouseCoopers Brasil*, 2018. (<https://www.pwc.com.br/pt/estudos/assets/2018/gecs-18.pdf>).
- Sarno, R., Dewandono, R. D., Ahmad, T., Naufal, M. F., and Sinaga, F. 2015. "Hybrid association rule learning and process mining for fraud detection," *IAENG International Journal of Computer Science* (42:2), pp. 59-72.
- Sarno, R., Sinaga, F. and Sungkono, K.R. 2020. "Anomaly detection in business processes using process mining and fuzzy association rule learning," *J Big Data* (7:1). (<https://doi.org/10.1186/s40537-019-0277-1>)
- Silva, F. A. R. 2018. "Analytical intelligence in processes: Data science for business," *IEEE Latin America Transactions* (16:8), pp. 2240-2247. (doi:10.1109/TLA.2018.8528241)
- Tavares, G. M., Turrisi Da Costa, V. G., Martins, V. E., Ceravolo, P., and Barbon, S. 2018. "Anomaly detection in business process based on data stream mining," in *Proceedings of the ACM International Conference Proceeding Series*, pp. 120-127. (doi:10.1145/3229345.3229362)
- Van Der Aalst, W. M. P., and De Medeiros, A. K. A. 2005. "Process mining and security: Detecting anomalous process executions and checking process conformance," *Electronic Notes in Theoretical Computer Science* (121:SPEC. ISS.), pp. 3-21. (doi:10.1016/j.entcs.2004.10.013)
- Van Der Aalst, W. M. P., Weijters, T., and Maruster, L. 2004. "Workflow mining: Discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering*, (16:9), pp. 1128-1142. (doi:10.1109/TKDE.2004.47)
- Yang, W.-S., and Hwang, S.-Y. 2006. "A process-mining framework for the detection of healthcare fraud and abuse," *Expert Systems with Applications* (31:1), pp. 56-68. (doi:10.1016/j.eswa.2005.09.003)