

# Evolução e tendências das Técnicas de Machine Learning aplicadas à Fiscalização Tributária: Uma Análise Bibliométrica

## Evolution and trends of Machine Learning Techniques applied to Tax Inspection: A Bibliometric Analysis

**Lucas Alvarenga** Bacharel em Engenharia Mecânica.  
<https://orcid.org/0009-0002-4178-4369> Universidade Católica de Brasília (UCB) -  
Brasil. [lucasalvarenga09@gmail.com](mailto:lucasalvarenga09@gmail.com)

**Eduardo Amadeu Dutra Moresi** Doutor em Ciência da Informação.  
Universidade Católica de Brasília (UCB) -  
<https://orcid.org/0000-0001-6058-3883> Brasil. [eduardo.moresi@gmail.com](mailto:eduardo.moresi@gmail.com)

**Edilson Ferneda** Doutor em Computação. Universidade Católica  
<https://orcid.org/0000-0003-4164-5828> de Brasília (UCB) - Brasil.  
[eferneda@gmail.com](mailto:eferneda@gmail.com).

**Fabrício Ziviani** Doutor em Ciência da Informação.  
<https://orcid.org/0000-0002-2705-846X> Universidade Católica de Brasília (UCB) -  
Brasil. [fazist@hotmail.com](mailto:fazist@hotmail.com)

**Matheus Silva de Paiva** Doutor em Economia. Universidade Católica  
<https://orcid.org/0000-0001-9882-1496> de Brasília (UCB) - Brasil.  
[matheus.paiva@p.ucb.br](mailto:matheus.paiva@p.ucb.br)

### RESUMO

Este artigo propõe uma abordagem bibliométrica para analisar a evolução e tendências sobre as técnicas de *machine learning* aplicadas à fiscalização tributária. O estudo se baseia em uma revisão bibliográfica de documentos científicos disponíveis na base Scopus. A metodologia adotada emprega ferramentas de análise bibliométrica implementadas pelo pacote Bibliometrix para extrair métricas bibliográficas e estatísticas descritivas. Além disso, o VOSviewer foi utilizado para a visualização da rede de coocorrência de palavras-chave do autor, proporcionando uma compreensão mais profunda das relações entre as diversas áreas do conhecimento e os trabalhos publicados. Para se obter algumas métricas da rede de coocorrência, utilizou-se o software Gephi. Também se fez uso da ferramenta Loglet Lab 4 para a análise temporal das publicações, permitindo identificar padrões de crescimento e projeção de saturação no interesse acadêmico sobre o tema ao longo do tempo. Os resultados obtidos revelam *insights* sobre as principais áreas de foco, contribuições mais influentes, colaborações entre pesquisadores e a dinâmica temporal da produção científica relacionada à aplicação de técnicas de Machine Learning na fiscalização tributária. Essa abordagem bibliométrica não apenas destaca o estado atual da pesquisa nesse campo, mas também oferece direcionamentos para futuros estudos e aplicações mais aprofundadas na área tributária.

**Palavras-chave:** machine learning; fiscalização tributária; análise bibliométrica.

#### ABSTRACT

This paper proposes a bibliometric approach to analyze the evolution and trends concerning machine learning techniques applied to tax auditing. The study is based on a bibliographic review of scientific documents available in the Scopus database. The adopted methodology employs bibliometric analysis tools implemented by the Bibliometrix package to extract bibliographic metrics and descriptive statistics. Furthermore, VOSviewer was used for the visualization of the author keyword co-occurrence network, providing a deeper understanding of the relationships between various knowledge areas and published works. The Loglet Lab 4 tool was also utilized for the temporal analysis of publications, enabling the identification of growth patterns and saturation projections in academic interest on the topic over time. The results obtained reveal insights into the main focus areas, most influential contributions, collaborations among researchers, and the temporal dynamics of scientific production related to the application of machine learning techniques in tax auditing. This bibliometric approach not only highlights the current state of research in this field but also offers directions for future studies and more in-depth applications in the tax area.

**Keywords:** machine learning; tax auditing; bibliometric analysis.

Recebido em 18/10/2024. Aprovado em 28/01/2025. Avaliado pelo sistema *double blind peer review*. Publicado conforme normas da APA.

<https://doi.org/10.22279/navus.v16.2037>

## 1 INTRODUÇÃO

Na última década, a tecnologia de Inteligência Artificial (IA) e *Machine Learning* (ML) viu avanços rápidos. Este progresso impulsionou transformações significativas em diversas esferas da sociedade, desde a indústria até a tomada de decisões governamentais (Bosdag, 2023).

O aumento do volume e complexidade das transações comerciais, a diversificação dos modelos de negócios e a globalização econômica têm desafiado as práticas tradicionais de fiscalização tributária (Adelakun et al., 2024). Diante desse cenário, as capacidades preditivas e analíticas do ML oferecem uma abordagem inovadora para aprimorar a eficácia e eficiência dos processos fiscais (Alsadhan, 2023). A capacidade dessas técnicas em lidar com grandes conjuntos de dados, identificar padrões complexos e automatizar tarefas analíticas despertou o interesse de pesquisadores e profissionais da administração tributária (Alsadhan, 2023).

Este estudo analisa a interseção entre os temas de ML e fiscalização tributária, oferecendo uma abordagem inovadora para compreender essa área em evolução. Ao investigar a produção científica nesse domínio, pretende-se identificar tendências emergentes, como a aplicação de algoritmos de aprendizado de máquina na detecção de irregularidades fiscais ou na otimização de sistemas de arrecadação. Além disso, por meio da análise de citações e cocitações, é possível identificar os trabalhos mais influentes nesse campo interdisciplinar.

A análise proposta permite identificar as áreas de foco e as contribuições mais influentes, mapear colaborações entre pesquisadores e entender a dinâmica temporal da produção científica. Além disso, ferramentas como VOSviewer<sup>1</sup> para visualizar a rede de coocorrência de palavras-chave, Gephi<sup>2</sup> para analisar medidas de rede e Loglet Lab 4<sup>3</sup> para análise temporal das publicações proporcionam uma compreensão mais profunda das relações entre áreas do conhecimento e a evolução do interesse acadêmico.

A justificativa para esta pesquisa é mapear o panorama atual da aplicação de técnicas de ML na fiscalização tributária. Compreender a evolução dessa interrelação contribui para consolidar o conhecimento nesse domínio e oferece direcionamentos para futuros estudos e aplicações para gestores públicos, pesquisadores e profissionais que buscam incorporar abordagens inovadoras na administração fiscal. A exploração da base de documentos do Scopus utilizou as ferramentas Bibliometrix<sup>4</sup>, VOSviewer, Gephi e Loglet Lab 4.

## 2 CONSTRUÇÃO DO CORPUS DE PESQUISA

### 2.1 Construção da expressão de busca

A construção do termo de busca foi um processo iterativo e estratégico, visando a abrangência e relevância dos documentos na base Scopus sobre a interseção entre "artificial intelligence" e "tax". Inicialmente, uma busca foi realizada com a expressão "artificial intelligence" AND "tax", resultando em 488 documentos.

As palavras-chave mais recorrentes da pesquisa inicial foram identificadas e incorporadas a uma nova expressão de busca expandida, utilizando o conector OR para cada termo adicional. A Tabela 1 apresenta a quantidade de documentos encontrados na base Scopus para cada etapa da pesquisa, indicando a acumulação daqueles localizados durante o processo de expansão da expressão de busca.

---

<sup>1</sup> <https://www.vosviewer.com/>

<sup>2</sup> <https://gephi.org/>

<sup>3</sup> <https://logletlab.com/>

<sup>4</sup> <https://www.bibliometrix.org/>

**Tabela 1** - Construção da expressão de busca

Expressão de Busca	Quantidade de Documentos
"artificial intelligence" AND tax	490
("artificial intelligence" AND tax) OR ("artificial intelligence" AND taxation)	561
("artificial intelligence" AND tax) OR ("artificial intelligence" AND taxation) OR ("artificial intelligence" AND gst)	594
("artificial intelligence" AND tax) OR ("artificial intelligence" AND taxation) OR ("artificial intelligence" AND gst) OR ("machine learning" AND tax)	946
("artificial intelligence" AND tax) OR ("artificial intelligence" AND taxation) OR ("artificial intelligence" AND gst) OR ("machine learning" AND tax) OR ("machine learning" AND taxation)	987
("artificial intelligence" AND tax) OR ("artificial intelligence" AND taxation) OR ("artificial intelligence" AND gst) OR ("machine learning" AND tax) OR ("machine learning" AND taxation) OR ("machine learning" AND gst)	1025

Fonte: Os autores (2024).

Esse método de construção do termo de busca ampliou a cobertura temática, incorporando variações e sinônimos relevantes. A quantidade acumulada de documentos representa a abrangência da pesquisa sobre a interseção entre IA e fiscalização tributária.

Após esta etapa, obteve-se um *corpus* com 1025 documentos. Nessa fase inicial exploratória, optou-se por não aplicar nenhum filtro específico, permitindo que o conjunto de documentos abrangesse uma ampla gama de informações da base de dados Scopus. A escolha da Scopus foi motivada por sua abrangência e reputação na comunidade acadêmica, oferecendo acesso a uma vasta coleção de artigos revisados por pares em diversas áreas. A ausência de filtro permitiu explorar a diversidade e a riqueza dos materiais disponíveis, garantindo uma análise abrangente e representativa para os objetivos do estudo.

## 2.2 Análise das áreas de conhecimento

A análise das áreas de publicação (Tabela 2) destaca a diversidade na produção científica. Ciência da Computação lidera com foco em tecnologia e IA. Engenharia investe em inovações práticas, enquanto Ciências Sociais estudam comportamento humano. Negócios e Contabilidade conectam academia e setor empresarial, e Matemática mantém interesse constante. Economia e Finanças analisam questões econômicas, e Ciências de Decisão tratam de métodos decisórios. Ciências Ambientais abordam preocupações ambientais, e Medicina tem relevância significativa. Física e Astronomia mantêm sua importância, enquanto Energia se concentra em questões energéticas. Outras áreas, como

Ciências dos Materiais, Ciências da Terra, Artes e Humanidades, Bioquímica, Genética, Ciências Agrárias e Biológicas, enriquecem ainda mais a diversidade científica, destacando a amplitude e interdisciplinaridade do conhecimento. Vale ressaltar que um documento pode pertencer a várias áreas, e os grupos não são excludentes.

**Tabela 2** - Análise das áreas de conhecimento

Áreas do Conhecimento	Quantidade de Documentos
Computer Science	579
Engineering	260
Social Sciences	202
Business, Management and Accounting	174
Mathematics	150
Economics, Econometrics and Finance	140
Decision Sciences	115
Environmental Science	68
Physics and Astronomy	65
Energy	61
Medicine	44
Materials Science	36
Earth and Planetary Sciences	32
Arts and Humanities	24
Biochemistry, Genetics and Molecular Biology	23
Agricultural and Biological Sciences	16
Multidisciplinary	12
Psychology	10
Chemistry	10
Neuroscience	10
Chemical Engineering	8
Pharmacology, Toxicology and Pharmaceutics	5
Health Professions	5
Immunology and Microbiology	1

Fonte: elaborada pelos autores (2024).

Considerando as áreas de fiscalização tributária, foi aplicado um filtro na busca para limitar o *corpus* às seguintes: *Computer Science*, *Engineering*, *Social Sciences*, *Business, Management and Accounting*, *Mathematics*, *Economics, Econometrics and Finance* e *Decision Sciences*. O resultado final foi 894 documentos. A consulta foi realizada em 15/11/2023 à base de dados da Scopus, podendo a quantidade variar devido à atualização do repositório.

### 3. ANÁLISE BIBLIOMÉTRICA DO CORPUS

#### 3.1 Documentos mais citados

Foi realizado um levantamento dos documentos mais citados no *corpus* final. O resultado está na Tabela 3.

**Tabela 3** - Documentos mais citados

Documento	Palavras-chave	Número de Citações
Kroll <i>et al.</i> (2017)	Sistemas de decisão automatizados; Justiça legal; Responsabilização algorítmica	403
Androutsopoulou <i>et al.</i> (2019)	Inteligência Artificial; Chatbots no setor público; Processamento de linguagem natural	220
Veale <i>et al.</i> (2018)	Responsabilização algorítmica; Viés algorítmico; Suporte à decisão; Policciamento preditivo;	218
Kontrimas e Verikas (2011)	Administração pública Avaliação em massa de imóveis; Perceptron multicamadas; Regressão de mínimos quadrados ordinários; Mapa auto-organizável; Regressão por vetores de suporte	122
Ephrati e Rosenschein (1991)	Inteligência artificial; Agentes autônomos; Agentes automatizados; Mecanismo de votação;	117
Wang (2011)	Tributação Inteligência artificial; Sistema especialista de design; Design de formulário; Teoria dos sistemas cinzentos; Engenharia Kansei; Ferramentas de máquinas; Regressão por vetores de suporte;	107
Paula <i>et al.</i> (2017)	Inteligência artificial; Crime; Lavagem de dinheiro; Sistemas de aprendizado; Tributação; Detecção de anomalias	99

Fonte: elaborada pelos autores.

Ao analisar os documentos mais citados, tem-se o seguinte panorama. Kroll *et al.* (2017) propuseram ferramentas para tornar os sistemas automatizados mais responsáveis, desafiando a ideia de que a transparência é a única solução. Androutsopoulou *et al.* (2019) discutiram o uso de IA no setor público, com foco em chatbots para aprimorar a comunicação entre governo e cidadãos. Veale *et al.* (2018) destacaram a importância da equidade em decisões públicas baseadas em algoritmos, apontaram desconexões entre instituições e pesquisas atuais, e sugeriram novas oportunidades de design. Kontrimas e Verikas (2011) compararam métodos de avaliação de propriedades, destacando abordagens de inteligência computacional que apresentaram bons resultados na detecção de zonas de valor. Ephrati e Rosenschein (1991) sugeriram o "imposto de Clarke" como mecanismo não manipulável para a coordenação entre agentes autônomos, reduzindo a necessidade de negociações explícitas.

Wang (2011) explorou a relação entre demandas dos clientes e formas de produtos usando teoria do sistema cinza e máquinas vetoriais de suporte. Ele apresentou um sistema especialista de design híbrido. Paula *et al.* (2016) abordaram fraudes em exportações usando aprendizado profundo não supervisionado para classificar exportadores brasileiros quanto à possibilidade de fraudes.

Esses textos oferecem perspectivas valiosas sobre a integração de tecnologia avançada em diversos campos. Eles destacam a necessidade de ética, transparência e responsabilidade na aplicação dessas inovações.

### 3.2 Documentos mais relevantes

A Scopus classifica os documentos por sua relevância com base nos argumentos de pesquisa no título, resumo e palavras-chave. A partir dessa ordenação, foi realizada uma nova análise nos documentos mais relevantes em relação ao termo de origem do corpus analisado (Tabela 4).

**Tabela 4** - Documentos mais relevantes

Documento	Palavras-chave	Número de citações
Kumar <i>et al.</i> (2023a)	Inteligência Artificial; Avaliação; Imposto sobre bens e serviços; Tributação	0
Shakil e Tasnia (2022)	Inteligência Artificial; Administração tributária; Evasão fiscal	2
Zadeh <i>et al.</i> (2023)	Inteligência artificial (IA); Big data; Questões éticas; Irã; Aprendizado de máquina; Imposto sobre a renda de aluguel;	0

	Imposto; Administração tributária; Evasão fiscal; Fraude fiscal	
Kumar <i>et al.</i> (2023b)	Inteligência Artificial; Avaliação; Imposto sobre bens e serviços; Tributação	0
Binder (2019)	Inteligência Artificial; Gestão de riscos; Evasão fiscal	2
Franic (2022)	Inteligência artificial; UE; Economia informal; Aprendizado de máquina; Evasão fiscal; Trabalho não declarado	1
Huang <i>et al.</i> (2022)	Inteligência artificial; Algoritmo SVM; Sistema tributário	0
Das e Kolya (2017)	Imposto sobre bens e serviços; Aprendizado de Máquina; Naive Bayes; Análise de Sentimento; Mineração de Texto	27
Phong <i>et al.</i> (2022)	Empresas; Aprendizado De Máquina; Risco Tributário	0
Raikov (2021)	Inteligência artificial; Modelagem cognitiva; Semântica cognitiva; Evasão fiscal; Planejamento tributário	4

Fonte: elaborada pelos autores (2024).

Kumar *et al.* (2023a) focaram no *Goods and Services Tax* (GST) com IA destacando transparência e competição na gestão tributária. Shakil e Tasnia (2022) exploraram a importância da IA na administração tributária na Ásia e no Pacífico, oferecendo recomendações para autoridades fiscais e corporações. Zadeh *et al.* (2023) estudaram o uso de *big data* contra evasão fiscal em aluguéis no Irã, utilizando sistemas de informação geográfica (SIG) e de Bancos de Dados de Grafos (GraphDB) para precinir evasão e fraudes na gestão tributária. Binder (2020) discutiu a Lei de Modernização da Tributação na Alemanha e o uso da IA para prevenir evasão fiscal. Kumar (2023) destacou o impacto do GST na Índia, enfatizando o papel da IA para identificar fraudes e simplificar as avaliações fiscais. Franic (2022) testou modelos de aprendizado de máquina em dados relacionados ao trabalho não declarado, para monitorar e confrontar evasão fiscal. Huang *et al.* (2022) analisaram a



integração da IA e gerenciamento de riscos fiscais para um sistema tributário inteligente. Das e Kolya (2017) exploraram análise de sentimento e mineração de opiniões em plataformas de mídia social sobre o GST na Índia. Phong et al. (2022) usaram aprendizado de máquina para prever riscos fiscais, recomendando *big data* e IA para agências fiscais. Raikov (2021) propôs combater a evasão fiscal corporativa com aprendizado profundo, redes neurais e modelagem cognitiva no diagnóstico de casos suspeitos na zona ártica da Rússia.

Os documentos revelam que os textos discutem IA, aprendizado de máquina e análise de dados na administração tributária. Evasão fiscal e detecção de fraudes são desafios centrais, sublinhando a importância da tecnologia nessas áreas. As mudanças legais, ferramentas estatísticas e modelos de aprendizado de máquina desempenham um papel crucial na modernização dos processos tributários e asseguram transparência. Perspectivas internacionais destacam a relevância global dessas abordagens. Análises de dados e modelos preditivos auxiliam na compreensão do comportamento dos contribuintes e na previsão de riscos fiscais. A mineração de opiniões em plataformas sociais e o uso de IA para aprimorar a gestão tributária também são discutidos. Esses textos mostram como a tecnologia está transformando a administração tributária em todo o mundo.

### 3.3 Análise de Coocorrência

A análise de coocorrência de palavras-chave é crucial na visualização e compreensão das inter-relações nos documentos científicos. Usando o VOSviewer, foi construída uma representação gráfica das coocorrências extraídas dos documentos da base Scopus.

Inicialmente, foram identificadas e pré-processadas as palavras-chave relevantes. Em seguida, construiu-se um dicionário de sinônimos (*thesaurus*) para melhorar a análise. Identificou-se 33 palavras-chave semelhantes a outro termo correlato e, em alguns casos, a forma do termo foi normalizada, como "machine-learning" para "machine learning". Isso possibilitou agrupar os termos do *thesaurus* para a contagem de coocorrências, melhorando a análise com o VOSviewer (van Eck & Waltman, 2023).

Na configuração dos parâmetros no VOSviewer, optou-se por um corte de ocorrências mínimas de palavras-chave de 5. Isso originou uma rede com 242 nós, 4739 arestas e 8 *clusters* identificados com as configurações padrões do software (Figura 1). Gerou-se uma rede de coocorrência, permitindo uma visualização interativa e ajustes dinâmicos, além de identificar a força das relações com base na frequência de coocorrência entre as palavras-chave.



Documentos	894
Taxa de Crescimento Anual %	13,42
Idade Média dos Documentos	5,19
Média de citações por documento	7,352
Referências	28145
CONTEÚDO DOS DOCUMENTOS	
Palavras-chave Plus (ID)	4588
Palavras-chave dos Autores (DE)	2332
AUTORES	
Autores	2249
Autores de documentos com autoria única	141
COLABORAÇÃO DOS AUTORES	
Documentos com autoria única	152
Coautores por Documento	2,94
% de colaborações internacionais	14,65
TIPOS DE DOCUMENTOS	
Artigo	347
Livro	10
Capítulo de livro	55
Artigo de conferência	408
Revisão de conferência	48
Editorial	3
Errata	3
Carta	1
Nota	1
Retratado	3
Revisão	15

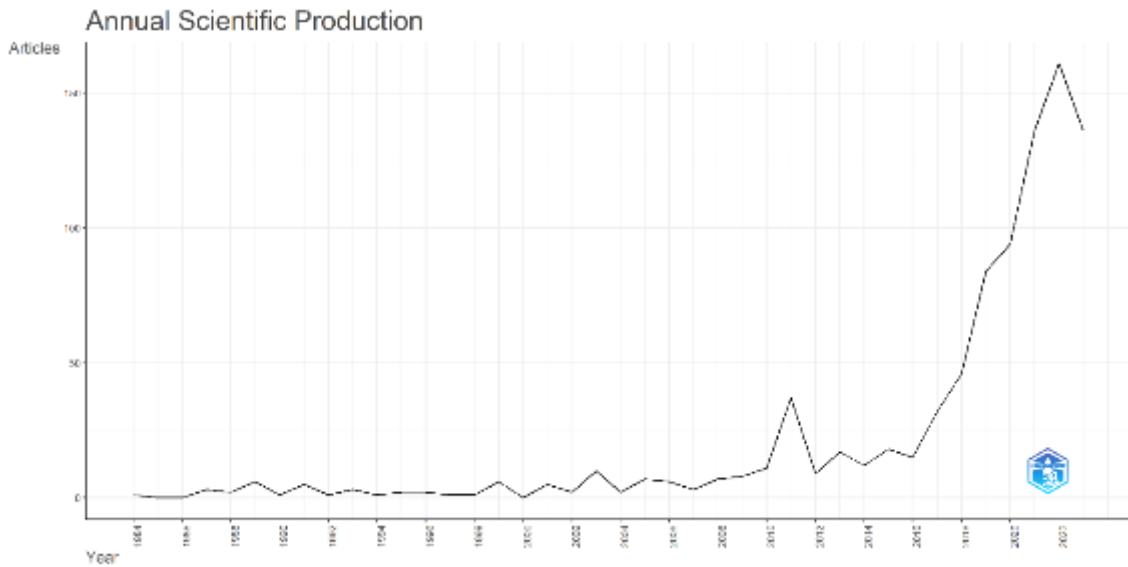
Fonte: gerada pelo Bibliometrix (2024).

Os dados de 1984 a 2024 incluem 894 documentos de 556 fontes, com uma média de 7,352 citações por documento, indicando impacto moderado. Eles são recentes, com média de idade de 5,19 anos. Há colaboração evidente, com 2,94 coautores por documento e 14,65% de colaborações internacionais. A maioria são artigos de conferências (408) e artigos (347). A pesquisa é colaborativa, internacional e focada em eventos técnico-científicos. Análises de palavras-chave e padrões de citação podem oferecer mais *insights*.

### 3.5 Evolução de documentos

A análise da produção acadêmica (Figura 2) mostra uma evolução notável ao longo das décadas. Inicialmente, havia poucos ou nenhum artigo, indicando menor interesse. A partir do fim dos anos 90, a publicação cresceu constantemente, tornando-se exponencial após 2010. Esse aumento se deve ao avanço tecnológico, especialmente em aprendizado de máquina e análise de *big data* na fiscalização tributária. Desde 2019, a produção teve um pico, atingindo 161 artigos em 2022. Com 136 em 2023, a tendência ascendente sugere uma atividade de pesquisa dinâmica e crescente.

**Figura 2 - Evolução de documentos**



Fonte: gerada pelo Bibliometrix (2024).

### 3.6 Fontes mais relevantes

O uso do Bibliometrix permitiu uma investigação abrangente das métricas bibliométricas das publicações, revelando a quantidade, qualidade e impacto das fontes. Nesta análise, explorou-se as principais métricas, como o número de documentos (NP); o índice H, G e M; número total de citações (TC); e ano inicial de publicação (PY), para identificar as mais influentes e impactantes (Tabela 6).

**Tabela 6 - Fontes mais relevantes**

Fonte	NP	h-index	TC	PY
Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	55	268	8	1998
International Conference on Artificial Intelligence, Management Science and Electronic Commerce, AIMSEC	28	9	2	2011
Advances in Intelligent Systems and Computing	26	43	4	2013
Lecture Notes in Networks and Systems	21	12	2	2020
International Conference on Artificial Intelligence and Law	16	258	9	1989
ACM International Conference Proceeding Series	14	55	3	2017
CEUR Workshop Proceedings	12	9	2	2009
Sustainability	11	251	6	2016
Smart Innovation, Systems and Technologies	9	1	1	2015
Communications in Computer and Information Science	8	26	2	2012

Fonte: gerada pelo Bibliometrix (2024).

A "LECTURE NOTES IN COMPUTER SCIENCE" lidera em artigos (55) e possui um h-index de 8, sugerindo impacto significativo, ao analisar a Tabela 6. A "INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, MANAGEMENT SCIENCE AND ELECTRONIC COMMERCE, AIMSEC" iniciou em 2011 com 28 artigos e um h-index de 2. "ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING" tem 26 artigos e h-index de 4, enquanto "LECTURE NOTES IN NETWORKS AND SYSTEMS" iniciou em 2020 com 21 artigos.

Outra fonte notável é "PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW", que possui 16 artigos e um h-index de 9, indicando forte influência na área desde 1989. "ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES" possui 14 artigos e um h-index de 3, enquanto "CEUR WORKSHOP PROCEEDINGS" iniciou em 2009 com 12 artigos e um h-index de 2.

"SUSTAINABILITY" começou em 2016, contribuindo com 11 artigos e apresentando um h-index de 6. "SMART INNOVATION, SYSTEMS AND TECHNOLOGIES" iniciou em 2015 com 9 artigos e um h-index de 1. "COMMUNICATIONS IN COMPUTER AND INFORMATION SCIENCE" possui 8 artigos e um h-index de 2 desde 2012.

Essas análises oferecem uma visão abrangente do impacto, produção acadêmica e longevidade dessas fontes, fornecendo uma visão mais precisa das fontes que desempenham um papel central na evolução do conhecimento nesse domínio em constante desenvolvimento.

### 3.7 Autores mais relevantes

Foi feito um levantamento com os autores destacados nas publicações nas áreas correlatas ao corpus analisado, assim como as fontes mais relevantes analisadas na seção anterior. Os resultados estão na Tabela 7.

**Tabela 7 - Autores mais relevantes**

Autor	NP	NP Frac	h-index	TC	PY
WANG C	5	1,36	5	193	2014
DONG B	7	1,22	4	37	2018
ZHENG Q	7	1,22	4	37	2018
WANG Y	7	1,38	3	14	2006
MASROM S	6	1,45	3	18	2019
RAHMAN RA	6	1,45	3	18	2019
ZHANG F	5	1,04	3	32	2010
ZHU X	5	1,12	3	26	2017
WANG S	4	1,78	3	37	2011
WU Y	4	0,6	3	24	2019

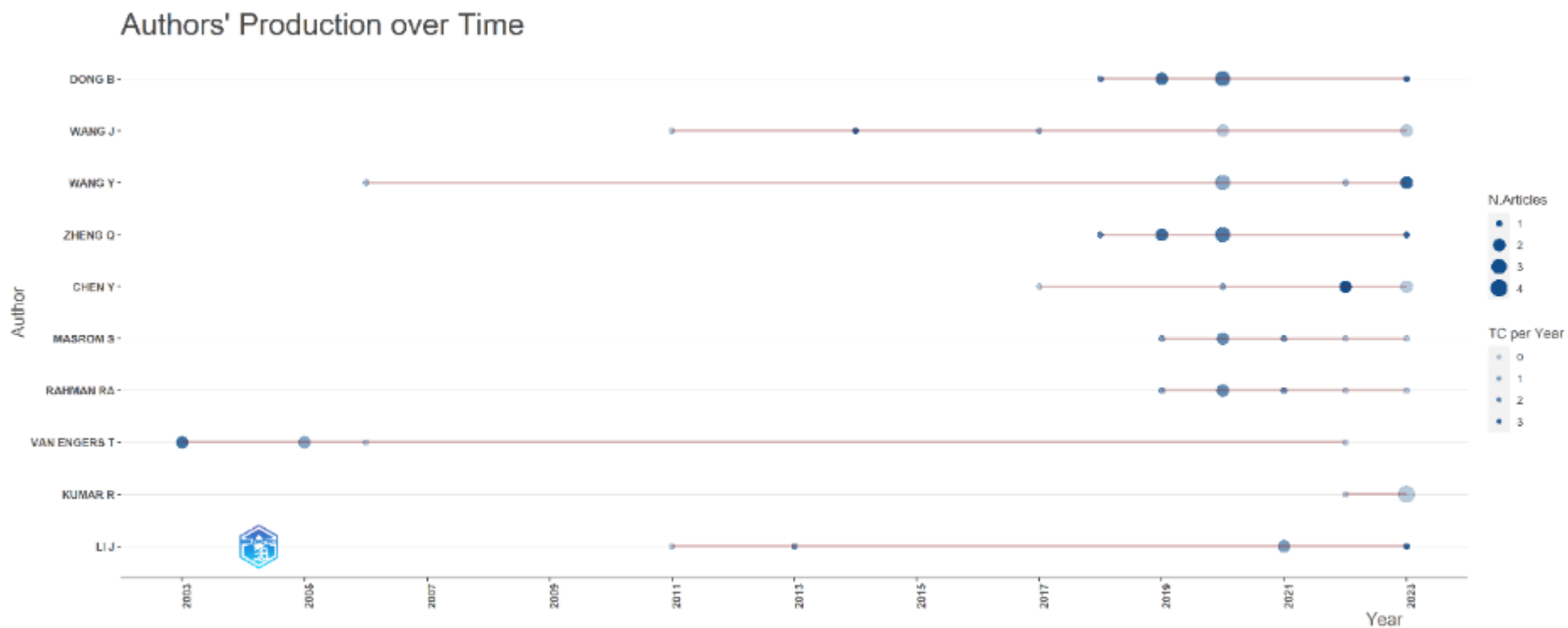
Fonte: gerada pelo Bibliometrix (2024).

As informações dos dez principais autores revelam diversas métricas de produção acadêmica. A análise revela que Wang C é o mais impactante, com 5 publicações, h-index de 5 e 193 citações, enquanto Dong B e Zheng Q têm o maior número de publicações (7), mas com impacto moderado (h-index de 4 e 37 citações). Wang S, com 4, destaca-se pelo impacto por artigo (NP Frac 1,78) e 37 citações. Autores como Masrom S, Rahman Ra, e Wu Y são mais recentes, com contribuições em ascensão. No geral, Wang C e Wang S têm maior influência, enquanto os demais estão em crescimento. Essas métricas oferecem *insights* sobre a influência, produtividade e impacto temporal de cada autor na academia.

### **3.8 Produção ao longo do tempo**

A análise da produção científica ao longo do tempo é fundamental para compreender a evolução do campo de estudo da aplicação de IA na fiscalização tributária. Nesta seção, explorou-se a trajetória temporal das publicações identificadas, utilizando o Bibliometrix. Elas se concentram a partir de 2019 (Figura 3). O autor VAN ENGERS T tem trabalhos de 2003 e citações ao longo dos anos. Outros com histórico de publicação mais antigo são WANG J, WANG Y e LI J.

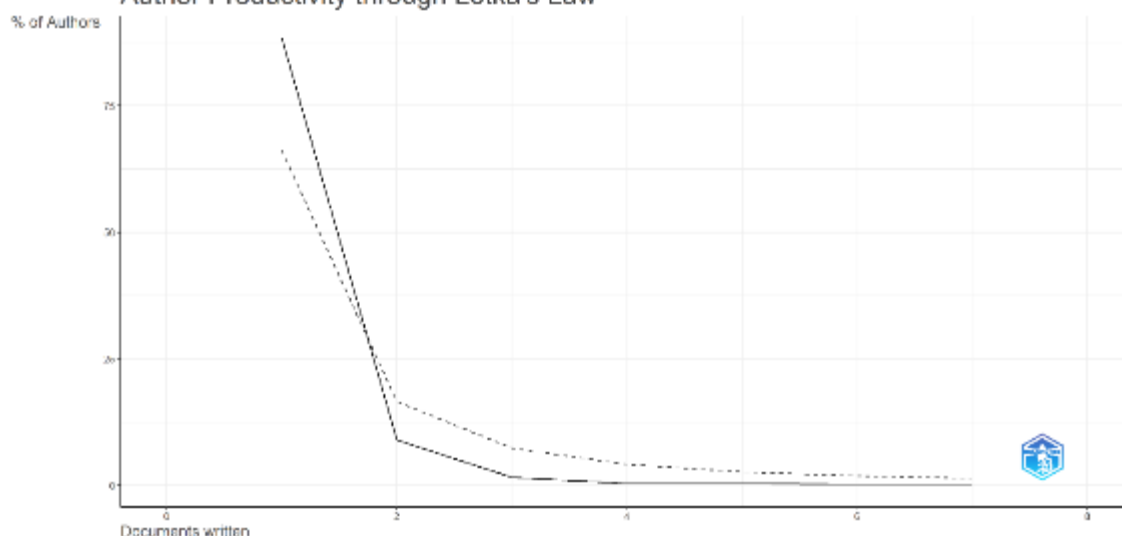
Figura 3 - Produção de documentos ao longo do tempo



Fonte: Gerada pelo Bibliometrix (2024).

A Lei de Lotka (Lotka, 1926) vem sendo usada para avaliar a produção acadêmica ao longo do tempo. Na bibliometria, refere-se a uma relação empírica que descreve a distribuição de produtividade dos autores em uma área de pesquisa. Essa lei sugere que um pequeno número de autores é responsável por uma grande parte da produção científica em uma área específica, enquanto a maioria contribui com uma quantidade pequena de trabalhos. A distribuição de autoria não é uniforme, mas segue uma distribuição de cauda longa, indicando desigualdade na produtividade científica. A Figura 4 ilustra o comportamento da Lei de Lotka. A maioria dos autores (88,4%) escreve apenas um documento, enquanto uma parcela menor produz dois (8,9%). Autores com mais de dois documentos correspondem a 2,7% do total.

**Figura 4** - Produção de documentos de acordo com a Lei de Lotka  
**Author Productivity through Lotka's Law**



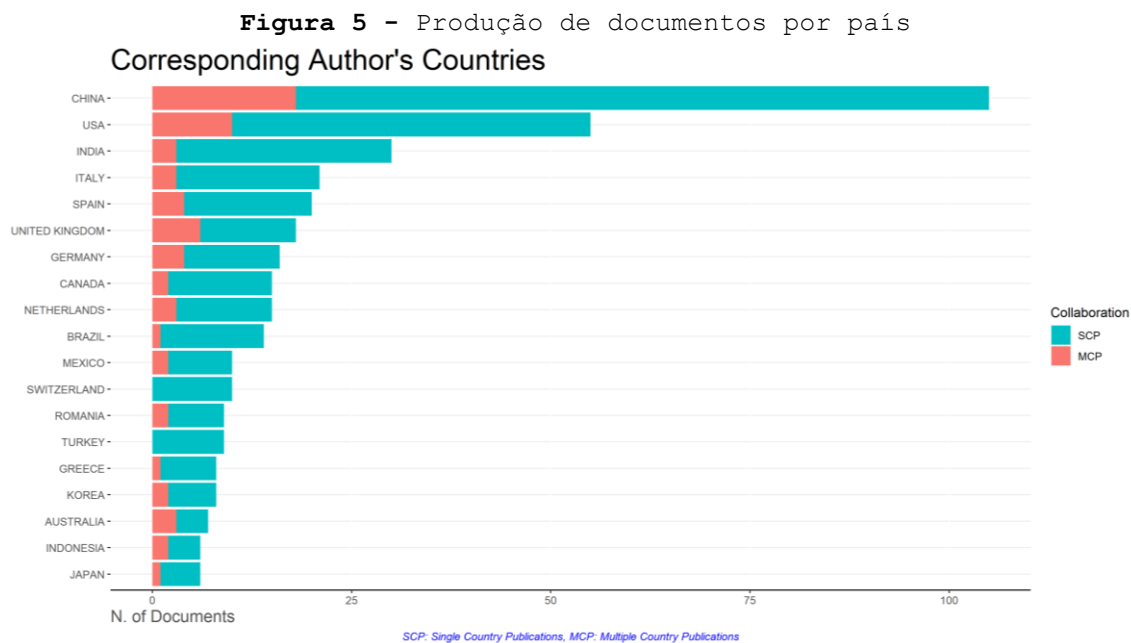
Fonte: gerada pelo Bibliometrix (2024).

### 3.9 Produção por país

A análise da produção científica por país é relevante para entender a distribuição global do conhecimento sobre a aplicação de IA na fiscalização tributária. Nesta seção, analisou-se a relação geográfica dos documentos do corpus, utilizando métricas e dados do Bibliometrix.

Dos 356 artigos analisados, 319 são de coautores do mesmo país (SCP - *Single Country Publication*). No entanto, a colaboração internacional (MCP - *Multiple Country Publications*) é notável, representando 37 artigos. A China lidera em MCP-Ratio, com 18 publicações resultantes de colaboração internacional, representando 17,1% do total. Os EUA seguem com 10 publicações, 18,2% do total. Isso destaca a importância da cooperação global. A China lidera com 105 artigos no tema de estudo, seguida pelos EUA com 55.





Ao investigar a produção científica por país, busca-se quantificar a participação de diferentes nações e entender como fatores regionais, políticos e econômicos influenciam as abordagens na interseção entre IA e fiscalização tributária. Essa análise geográfica é relevante para reconhecer as ênfases e contribuições de diferentes contextos nacionais, enriquecendo a compreensão da pesquisa nesse campo global.

### 3.10 Documentos com mais citações locais

A análise de citações locais (Tabela 8) dos documentos do corpus representa a influência mútua da interconectividade dos trabalhos na aplicação de IA na fiscalização tributária. Ao explorar as citações locais, busca-se identificar padrões de referências cruzadas entre os documentos, destacando as contribuições mais influentes e as discussões, fundamentais para o desenvolvimento do conhecimento nessa interseção temática.

**Tabela 8 - Documentos com mais citações locais**

Documento	LC	GC	LC/GC (%)	NLC	NGC
Liu <i>et al.</i> (2010)	5	15	33,33	11	3
Oberson (2019)	4	9	44,44	39	0,71
Zhu <i>et al.</i> 2018	4	12	33,33	45	0,81
Richins <i>et al.</i> 2020	2	12	16,67	25,43	1,77
Placencia <i>et al.</i> (2020)	2	2	100	25,43	0,29
Wu (2019)	2	7	28,57	19,5	0,56
Di Oliveira <i>et al.</i> (2021)	1	1	100	63,5	0,2
Jiang e Duan (2021)	1	2	50	63,5	0,41
Zhang (2020)	1	3	33,33	12,71	0,44
Mehta <i>et al.</i> (2020)	1	3	33,33	12,71	0,44

Fonte: gerada pelo Bibliometrix (2025).

A Tabela 8 mostra a influência de trabalhos específicos no corpus analisado. O artigo "Application of hierarchical clustering in tax inspection case-selecting" (Liu, 2010) tem forte presença interna, com 5 citações locais

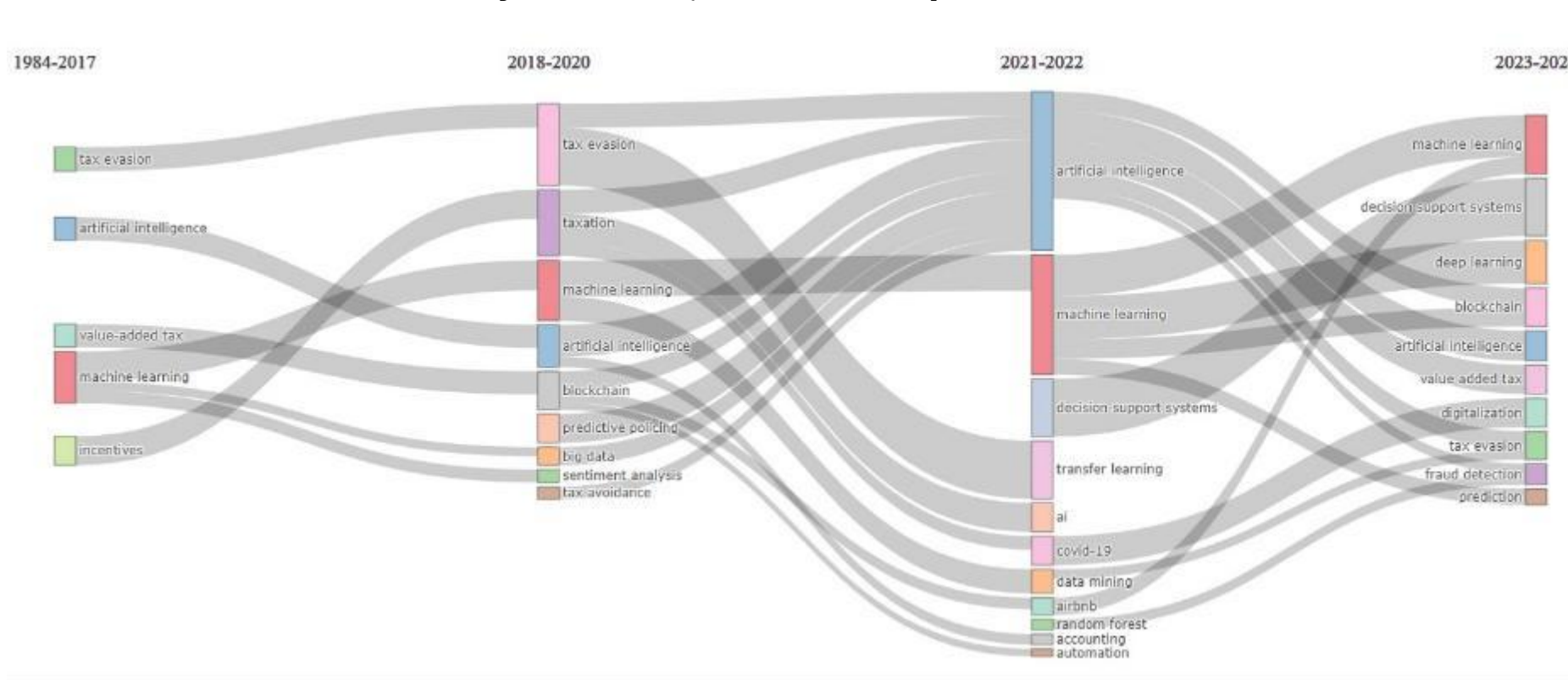
em 15 globais. "Taxing Robots: Helping the Economy to Adapt to the Use of Artificial Intelligence" (Oberson, 2019) também é relevante, com 4 citações locais em 9 globais. "IRTED-TL: An Inter-Region Tax Evasion Detection Method Based on Transfer Learning" (Zhu, 2018) destaca-se com 4 citações locais em 12 globais. Esses documentos moldaram a pesquisa na interseção entre IA e fiscalização tributária, conforme indicado pelas citações.

### **3.11 Evolução das palavras-chave**

A análise da evolução dos termos utilizados como palavras-chave pelos autores em cada documento científico é essencial para compreender a dinâmica e a evolução conceitual no campo da interseção entre IA e Fiscalização Tributária. Nesta seção, explorou-se como as palavras-chave ao longo do tempo refletem as mudanças e tendências no discurso acadêmico, destacando o surgimento de novos conceitos.

A análise da evolução dos termos realizada (Figura 6) revela tendências importantes. A IA manteve uma presença constante, aumentando sua relevância de 1984-2017 para 2018-2020 e permanecendo forte em 2021-2022. O termo "Machine Learning" expandiu-se, incluindo "Data Mining" e "Sentiment Analysis". "Big Data" e "Blockchain" se destacaram, indicando um interesse crescente. "Tax Evasion" permaneceu constante, enquanto "Taxation" teve um aumento notável. Tópicos emergentes incluem "Robot Tax", "Mass Appraisal", "Public Policy", "Security" e "Transfer Learning". O Stability Index sugere mudanças frequentes nos tópicos discutidos. Essa análise reflete a evolução dinâmica e diversificação das discussões em IA e áreas correlatas.

Figura 6 - Evolução dos termos de palavras-chave

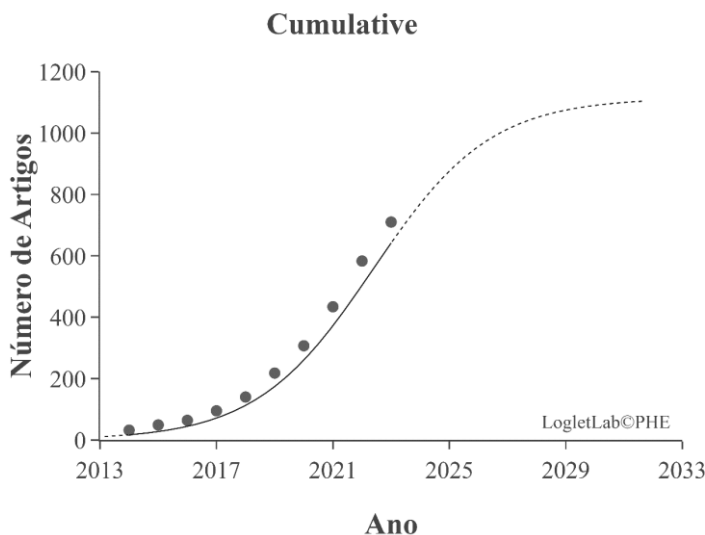


Fonte: Gerada pelo Bibliometrix (2024).

### 3.12 Ciclo de vida do tema

Na análise do ciclo de vida do tema, usou-se a ferramenta Loglet Lab 4 (BURG *et al.*, 2024) para mapear o comportamento temporal e modelar a trajetória do campo de estudo. Essa ferramenta permitiu uma análise mais aprofundada, aproximando o comportamento da produção científica a uma regressão logística, caracterizada por uma curva em forma de S (Figura 7).

**Figura 7** - Ciclo de vida do tema.



Fonte: Gerada pelo Loglet Lab 4 (2024).

Usando uma aproximação da publicação por uma regressão logística em 'S', observa-se que o tema deste trabalho está em crescimento. Considerando o ritmo atual, a saturação das publicações científicas deve ocorrer por volta de 2033.

Ao aplicar uma abordagem de regressão logística, constata-se que o campo de estudo deste trabalho ainda está em crescimento. O acompanhamento do ritmo de evolução sugere a expectativa de atingir um ponto de saturação das publicações científicas em 2033. Essa projeção fornece uma perspectiva sobre a trajetória futura do tema, indicando um cenário onde o volume de publicações poderá estabilizar-se após um período de crescimento.

### 3.13 Tendências das técnicas de machine learning aplicadas à fiscalização tributária

As técnicas de ML têm desempenhado um papel significativo na modernização da fiscalização tributária, com destaque para a detecção de fraudes e evasão fiscal. Algoritmos avançados, como aprendizado profundo e análise de anomalias, têm sido amplamente utilizados para identificar padrões fraudulentos (Alsadhan, 2023; Paula *et al.*, 2016). Além disso, a análise preditiva e a automação dos processos fiscais são outras tendências importantes, com redes neurais e métodos de regressão por vetores de suporte sendo aplicados para prever comportamentos fiscais e otimizar auditorias (Wang, 2011; Zhu *et al.*, 2018).

A interoperabilidade entre ML e tecnologias emergentes, como *big data* e blockchain, também tem mostrado grande potencial na melhoria da eficiência e transparência dos sistemas fiscais. Essas tecnologias ajudam a gerenciar e validar transações em tempo real, conforme observado por Zadeh *et al.* (2023). Paralelamente, técnicas de processamento de linguagem natural estão sendo utilizadas para personalizar e otimizar políticas fiscais, proporcionando uma melhor compreensão das necessidades dos contribuintes (Androutsopoulou *et al.*, 2019).

---

A análise da rede de coocorrência de palavras-chave revelou *clusters* temáticos em áreas como detecção de fraudes, evasão fiscal e políticas públicas, refletindo o interesse crescente em soluções tecnológicas para melhorar a conformidade fiscal e reduzir irregularidades. O impacto global dessas tendências é evidente na presença significativa de publicações provenientes de países como China e EUA, com cerca de 14,65% das pesquisas resultando de colaborações internacionais, o que sublinha a relevância universal do tema.

A produção acadêmica na interseção de ML e fiscalização tributária cresceu significativamente após 2010, atingindo um pico em 2022, impulsionada pelos avanços em ML e *big data*. Projeções indicam que a pesquisa deve atingir um ponto de saturação por volta de 2033, com a consolidação dessas aplicações no setor tributário. Esse panorama destaca a importância do uso de ML para promover eficiência (Alsadhan, 2023; Paula et al., 2016; Wang, 2011; Zhu et al., 2018), conformidade e inovação nos sistemas fiscais (Androutsopoulou et al., 2019, Zadeh et al., 2023).

#### 4. CONCLUSÃO

Este estudo realizou uma análise bibliométrica sobre o uso de técnicas de ML na fiscalização tributária, buscando compreender a evolução e os padrões dessa interrelação ao longo do tempo. Usando Bibliometrix, VOSviewer, Gephi e Loglet Lab 4, explorou-se a base de documentos extraída de publicações científicas Scopus, conduzindo uma investigação estruturada em diferentes asp

O estudo começou com a construção do *corpus* de pesquisa usando a base de dados Scopus e delineando os parâmetros para a análise. A seleção dos documentos formou uma base sólida, assegurando a representatividade necessária para entender a pesquisa sobre a aplicação de técnicas de ML na fiscalização tributária.

Em sequência, realizou-se uma análise exploratória, identificando padrões e tendências na literatura. Essa etapa foi relevante para estabelecer os pontos de convergência e divergência para refinar a base de documentos e orientar futuras análises.

Os documentos mais citados e relevantes delinearão os pilares do conhecimento consolidado nesse campo, validando contribuições essenciais e apontando direções de aprofundamento. A análise de coocorrência revelou padrões temáticos emergentes, evidenciando a complexidade e interconexão dos conceitos abordados na pesquisa.

Ao explorar informações gerais do *corpus*, como a evolução temporal, fontes, autores, produção por país e documentos mais citados, estabeleceu-se uma compreensão holística da dinâmica da pesquisa. A trajetória do tema, desde seus estágios iniciais até as tendências contemporâneas, contextualizou o presente e ofereceu perspectivas para o desenvolvimento do assunto.

A análise bibliométrica desta pesquisa sobre a aplicação de técnicas de ML na fiscalização tributária proporcionou uma visão abrangente do estado atual da interseção entre esses domínios. Destaca-se a importância de explorar as inter-relações e considerar outros repositórios de publicações científicas. O dinamismo das interações científicas oferece um campo vasto e promissor para futuras pesquisas e descobertas.

#### Referências

Adelakun, B. O., Nembe, J. K., Oguejiofor, B. B., Akpuokwe, C. U., & Bakare, S. S. (2024). Legal frameworks and tax compliance in the digital economy: a finance perspective. *Engineering Science & Technology Journal*, 5(3), 844-853. <https://doi.org/10.51594/estj.v5i3.922>

- 
- Alsadhan, N. (2023). A Multi-Module Machine Learning Approach to Detect Tax Fraud. *Computer Systems: Science & Engineering*, 46(1), 241-253. <https://doi.org/10.32604/csse.2023.033375>
- Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, 36(2), 358-367. <https://doi.org/10.1016/j.giq.2018.10.001>
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Binder, N. B. (2020). Artificial intelligence and taxation: risk management in fully automated taxation procedures. In: T. Wischmeyer & T. Rademacher (Eds.). *Regulating Artificial Intelligence* (pp. 295-306), Springer. [https://doi.org/10.1007/978-3-030-32361-5\\_13](https://doi.org/10.1007/978-3-030-32361-5_13)
- Bozdog, A. A. (2023). AIsmosis and the pas de deux of human-AI interaction: Exploring the communicative dance between society and artificial intelligence. *Online Journal of Communication and Media Technologies*, 13(4). <https://doi.org/10.30935/ojcm/13414>
- Burg, D., Schachter, E., Meyer, P., Yung, J., Wernick, I., & CURRY, A. (2017). Loglet Lab. Versão 4.0. <http://logletlab.com>.
- Das, S., & Kolya, A. K. (2017). Sense GST: Text mining & sentiment analysis of gst tweets by naive bayes algorithm. *Proceedings of the 3<sup>rd</sup> International Conference on Research in Computational Intelligence and Communication Networks*, p. 239-244. <https://doi.org/10.1109/ICRCICN.2017.8234513>
- Di Oliveira, V., Chaim, R. M., Weigang, L.; Bittencourt Neto, S. A. P., & Rocha Filho, G. P. (2021). Towards a Smart Identification of Tax Default Risk with Machine Learning. *Proceedings of the 17<sup>th</sup> International Conference on Web Information Systems and Technologies*, Volume 1, 422-429. <https://doi.org/10.5220/0010712200003058>
- Ephrati, E., & Rosenschein, J. S. (1991). The clarke tax as a consensus mechanism among automated agents. *Proceedings of the 9<sup>th</sup> National Conference on ARTIFICIAL intelligence*, Vol. 1, 173-178. <https://cdn.aaai.org/AAAI/1991/AAAI91-028.pdf>
- Franic, J. (2024). What do we really know about the drivers of undeclared work? an evaluation of the current state of affairs using machine learning. *AI & society*, 39, 597-616. <https://doi.org/10.1007/s00146-022-01490-3>
- Huang, W., He, L., & Zhang, J. (2022). Artificial intelligence technology and tax risk management innovation. *Proceedings of the International Conference on Computer, Artificial Intelligence, and Control Engineering*, 362-366. <https://doi.org/10.1117/12.2641091>
- Jiang, C., & Duan, H. (2021). Research and implementation of Intelligent Service Platform for Flexible Employment in Internet Sharing Economy. *Proceedings of the IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology*, 237-241. <https://doi.org/10.1109/CEI52496.2021.9574501>
- Kontrimas, V., & Verikas, A. (2011). The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing*, 11(1), 443-448. <https://doi.org/10.1016/j.asoc.2009.12.003>

---

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633-705.  
[https://scholarship.law.upenn.edu/penn\\_law\\_review/vol165/iss3/3](https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3)

Kumar, R., Malholtra, R. K., Singh, R., Kathuria, S., Balyan, R., & Pal, P. (2023a). Artificial intelligence role in electronic invoice under goods and services tax. *Proceedings of the International Conference on Computational Intelligence, Communication Technology and Networking*, 140-143.  
<https://doi.org/10.1109/CICTN57981.2023.10140870>

Kumar, R., Malholtra, R. K., Pandey, S., Gehlot, A., Gautam, I., & Chamola, S. (2023b). Role of artificial intelligence in input tax credit reconciliation. *Proceedings of the 3<sup>rd</sup> International Conference on Pervasive Computing and Social Networking*, 497-501.  
<https://doi.org/10.1109/ICPCSN58827.2023.00086>

Liu, X., Pan, D., Chen, S. (2010). Application of hierarchical clustering in tax inspection case-selecting. *Proceedings of the International Conference on Computational Intelligence and Software Engineering*.  
<https://doi.org/10.1109/CISE.2010.5676711>

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317-323.

Mehta, D., Desai, D., & Pradeep, J. (2020). Machine learning fund categorizations. *Proceeding of the 1<sup>st</sup> ACM International Conference on AI in Finance*. <https://doi.org/10.1145/3383455.3422555>

Oberson, X. (2019). *Taxing Robots: Helping the Economy to Adapt to the Use of Artificial Intelligence*. Elgaronline.  
<https://doi.org/10.4337/9781788976527>

Paula, E. L., Ladeira, M., Carvalho, R. N., & Marzagão, T. (2016). Deep learning anomaly detection as support fraud investigation in brazilian exports and anti-money laundering. *Proceedings of the 15<sup>th</sup> International Conference on Machine Learning and Applications*, 954-960.  
<https://doi.org/10.1109/ICMLA.2016.0172>

Phong, N. A., Tam, P. H., & Cuong, L. Q. (2022). Forecasting tax risk by machine learning: Case of firms in ho chi minh city. In: A. J. Tallón-Ballesteros (Ed.). *Fuzzy Systems and Data Mining VIII* (pp. 66-71). IOS Press. <https://doi.org/10.3233/FAIA220371>

Placencia, J., Hallo, M., & Lujan-Mora, S. (2020). Detection of Taxpayers with High Probability of Non-payment: An Implementation of a Data Mining Framework. *Proceedings of the 15<sup>th</sup> Iberian Conference on Information Systems and Technologies*. <https://doi.org/10.23919/CISTI49556.2020.9140837>

Raikov, A. (2021). Decreasing tax evasion by artificial intelligence. *IFAC-PapersOnLine*, 54(13), 172-177. <https://doi.org/10.1016/j.ifacol.2021.10.440>

Richins, D., Doshi, D., Blackmore, M., Nair, A. T., Pathapati, N., Patel, A., Daguman, B., Dobrijalowski, D., Illikkal, R., Long, K., Zimmerman, D., Reddi, V. J. (2020). Missing the forest for the trees: End-to-end ai application performance in edge data centers. *Proceedings of the IEEE International Symposium on High Performance Computer Architecture*, 515-528.  
<https://doi.org/10.1109/HPCA47549.2020.00049>

Shakil, M. H.; Tasnia, M. Artificial intelligence and tax administration in asia and the pacific. In: HENDRIYETTY, N.; EVANS, C.; KIM, C. J.

---

Taghizadeh-Hesary, F. (Eds). Taxation in the Digital Economy. Routledge, 2022. p. 45-55. <https://doi.org/10.4324/9781003196020-4>

Van Eck, N. J., & Waltman, L. (2023). *Manual for VOSviewer version, v. 1.6.20*. [https://www.vosviewer.com/documentation/Manual\\_VOSviewer\\_1.6.20.pdf](https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf)

Veale, M., Kleek, M. V., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. *Proceedings of the Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3173574.3174014>

Wang, K.-C. (2011). A hybrid kansei engineering design expert system based on grey system theory and support vector regression. *Expert Systems with Applications*, 38(7), 8738-8750. <https://doi.org/10.1016/j.eswa.2011.01.083>

Wu, Y., Zheng, Q., Gao, Y., Dong, B., Wei, R., Zhang, F., & He, H. (2019). TEDM-PU: A Tax Evasion Detection Method Based on Positive and Unlabeled Learning. *Proceedings of the IEEE International Conference on Big Data*, 1681-1686. <https://doi.org/10.1109/BigData47090.2019.9006325>

Zadeh, S. A., Iwendi, C., Uhumuavbi, I., & Boulouard, Z. (2023). A New AI-Based Approach for Rental Tax Evasion Management in Iran (Ethical Consideration). *Lecture Notes in Networks and Systems*, 735, 451-468. [https://doi.org/10.1007/978-3-031-37164-6\\_34](https://doi.org/10.1007/978-3-031-37164-6_34)

Zhang, M. Practical thinking on the new tax service in the era of artificial intelligence. *Proceedings of the International Conference on E-Commerce and Internet Technology*, 201-203. <https://doi.org/10.1109/ECIT50008.2020.00052>

Zhu, X., Yan, Z., Ruan, J., Zheng, Q., & Dong, B. (2018). IRTED-TL: An Inter-Region Tax Evasion Detection Method Based on Transfer Learning. *Proceedings of the 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering*, 1224-1235. <https://doi.org/10.1109/TrustCom/BigDataSE.2018.00169>