



**MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO DELTA DO PARNAÍBA – UFDPAr
CAMPUS MINISTRO REIS VELLOSO
COORDENAÇÃO DO CURSO DE CIÊNCIAS ECONÔMICAS**

BRUNO CASTRO ALVES

**MAPEAMENTO BIBLIOMÉTRICO DA EVOLUÇÃO DO *MACHINE LEARNING* NA
ECONOMIA**

**Parnaíba - Piauí
2025**

BRUNO CASTRO ALVES

**MAPEAMENTO BIBLIOMÉTRICO DA EVOLUÇÃO DO *MACHINE LEARNING* NA
ECONOMIA**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciências Econômicas da UFDPAr - Universidade Federal do Delta do Parnaíba, como requisito parcial para obtenção do título de bacharel em Ciências Econômicas.

Orientador(a): Prof. Dr. Fábio Júnior Clemente Gama

**Parnaíba - Piauí
2025**

FICHA CATALOGRÁFICA
Universidade Federal do Delta do Parnaíba

A474m Alves, Bruno Castro

Mapeamento bibliométrico da evolução do *machine learning* na economia [recurso eletrônico] / Bruno Castro Alves. – 2025.

56 p.

TCC (Bacharelado em Ciências Econômicas) – Universidade Federal do Delta do Parnaíba, 2025.

Orientação: Prof. Dr. Fábio Júnior Clemente Gama.

1. Ciência de dado. 2. Aprendizado de Máquina. 3. Bibliometria. 4. Economia. I. Gama, Fábio Júnior Clemente. II. Título.

CDD: 330

BRUNO CASTRO ALVES

**MAPEAMENTO BIBLIOMÉTRICO DA EVOLUÇÃO DO *MACHINE LEARNING* NA
ECONOMIA**

Monografia apresentada como exigência para a obtenção do título de Bacharel em Ciências Econômicas pela Universidade Federal do Delta do Parnaíba (UFDPAr), Campus Ministro Reis Velloso, submetida à aprovação da banca examinadora composta pelos seguintes membros:

BANCA EXAMINADORA

Prof. Dr. Fábio Júnior Clemente Gama
(Orientador) – UFDPAr

Prof. Dr. José Natanael Fontenele de Carvalho
(Avaliador) – UFDPAr

Consultor Igor Machado Torres
(Avaliador) – UFJF

AGRADECIMENTOS

Dedico este trabalho, primeiramente, a Deus, por guiar cada passo da minha caminhada e por me conceder força nos momentos mais desafiadores. À minha namorada, Rita de Cássia, pelo apoio incondicional, pela paciência e por estar sempre ao meu lado em todas as etapas desta jornada.

À minha família, pelo amor, incentivo e suporte financeiro indispensáveis. Em especial, agradeço à Márcia Maria, Maria Isolda, Raimundo Nonato, Marcilene Krause e Francisco Márcio, que sempre acreditaram no meu potencial. À Professora Fátima, pelo acolhimento e pela oportunidade de participar do Projeto Laços de Solidariedade e do Programa Entrelaços, experiências que enriqueceram minha formação.

À banca avaliadora, pelo tempo, leitura atenta e contribuições ao meu trabalho. À Coordenação do curso, na pessoa do Professor Natanael, pelo comprometimento com a formação dos alunos. Aos Professores Fábio e Caio, pelas valiosas orientações e pelo trabalho desenvolvido no Núcleo Nupace.

Aos meus amigos de turma, pelo companheirismo, incentivo e por serem parte essencial da minha motivação diária. A todos os professores do curso, pela dedicação e contribuição à minha formação. À Kaylane Manuele, pela valiosa contribuição na pesquisa relacionada e pelo incentivo constante que me ajudou a seguir em frente.

Em especial, à Professora Vera Barcelar, por plantar em mim o desejo de florescer em novos jardins, e ao Professor Diego Pacheco, por sua influência na minha formação ética e profissional, além do apoio constante ao longo do curso. A todos que, de alguma forma, contribuíram para a construção desse trabalho: gratidão!

"Entrega o teu caminho ao Senhor; confia nele, e ele o fará. E ele fará sobressair a tua justiça como a luz, e o teu juízo como o meio-dia. Descansa no Senhor, e espera nele; não te indignes por causa daquele que prospera em seu caminho, por causa do homem que executa astutos intentos. Deixa a ira, e abandona o furor; não te indignes de forma alguma para fazer o mal. Porque os malfeitores serão desarraigados; mas aqueles que esperam no Senhor herdarão a terra."

— Salmos 37:5–9 (SAGRADA, 2001)

RESUMO

Este trabalho mapeia a produção científica sobre a aplicação de Machine Learning (ML) na Economia por meio de uma análise bibliométrica. O estudo objetiva identificar a estrutura intelectual e social deste campo de pesquisa emergente. Utilizando o pacote *Bibliometrix* do software *R*, a pesquisa de caráter descritivo analisou 913 artigos únicos extraídos das bases de dados *Scopus* e *Web of Science*, publicados entre 1996 e 2024. Os resultados confirmam um crescimento exponencial nas publicações, com uma taxa anual de 21,42%, concentrado especialmente após 2017. A análise temática revelou um campo estruturado em um núcleo metodológico que alimenta duas frentes principais: uma consolidada em Finanças Quantitativas e outra, emergente, em Economia da Sustentabilidade. Adicionalmente, o estudo identificou uma liderança de publicações de instituições dos Estados Unidos e uma rede de colaboração global ainda fragmentada.

Palavras-chave: Ciência de Dados; Aprendizado de Máquina; Bibliometria; Economia.

ABSTRACT

This paper maps the scientific production on the application of *Machine Learning* (ML) in Economics through a bibliometric analysis. The study aims to identify the intellectual and social structure of this emerging research field. Using the *Bibliometrix* package from the R software, the descriptive research analyzed 913 unique articles extracted from the *Scopus* and *Web of Science* databases, published between 1996 and 2024. The results confirm an exponential growth in publications, with an annual growth rate of 21.42%, particularly concentrated after 2017. The thematic analysis revealed a field structured around a methodological core that feeds into two main fronts: one consolidated in Quantitative Finance and another, emerging, in the Economics of Sustainability. Additionally, the study identified a leadership in publications from institutions in the United States and a still fragmented global collaboration network.

Keywords: Data Science; Machine Learning; Bibliometrics; Economic Sciences.

LISTA DE FIGURAS

FIGURA 1 – Fluxograma metodológico	9
FIGURA 2 – Painel Geral	14
FIGURA 3 – Rede de colaboração internacional entre autores em produções científicas	23
FIGURA 4 – Rede de colaboração internacional entre países em produções científicas	32
FIGURA 5 – Evolução temática	34

LISTA DE GRÁFICOS

GRÁFICO 1 – Evolução das publicações ao longo do tempo	16
GRÁFICO 2 – Publicação dos periódicos ao longo do tempo	18
GRÁFICO 3 – Principais autores ao longo do tempo	22
GRÁFICO 4 – Principais tópicos emergentes	28
GRÁFICO 5 – Mapa temático	29
GRÁFICO 6 – Países mais citados	31

LISTA DE TABELAS

TABELA 1 –	Principais Periódicos que publicam artigos de <i>Machine Learning</i> Aplicado à Economia	17
TABELA 2 –	Principais Instituições em Pesquisa de <i>Machine Learning</i> na Economia	19
TABELA 3 –	Autores mais impactantes em artigos que versam sobre <i>Machine Learning</i> aplicado à Economia	21
TABELA 4 –	Papers mais citados em artigos de economia que utilizam ML .	25

LISTA DE QUADROS

QUADRO 1 – Descrição das variáveis analisadas	11
---	----

LISTA DE SIGLAS

AUROC	Característica de Operação do Receptor
ATE	Identificador de Objeto Digital
DOI	Identificador de Objeto Digital
IA	Inteligência Artificial
ML	Machine Learning
NBER	National Bureau of Economic Research
PIB	Produto Interno Bruto
PLN	Processamento de Linguagem Natural
WoS	Web of Science

SUMÁRIO

1 INTRODUÇÃO	1
2 MACHINE LEARNING EM ECONOMIA: UMA NOVA FRONTEIRA METODOLÓGICA	4
2.1 A REVOLUÇÃO NA INFERÊNCIA CAUSAL E NA AVALIAÇÃO DE POLÍTICAS PÚBLICAS	4
2.2 APLICAÇÕES EM MACROECONOMIA E ANÁLISE DE RISCO FINANCEIRO	5
2.3 FRONTEIRAS, DESAFIOS E UMA VISÃO CRÍTICA	6
3 PROCEDIMENTOS METODOLÓGICOS	8
3.1 COLETA E DELIMITAÇÃO DA AMOSTRA	9
3.2 TRATAMENTO DOS DADOS E ANÁLISE ESTRUTURAL	11
4 RESULTADOS	14
5 CONSIDERAÇÕES FINAIS	36
REFERÊNCIAS	39
ÍNDICE REMISSIVO	42

1 INTRODUÇÃO

Nas últimas décadas, o avanço das tecnologias de informação e comunicação, especialmente impulsionado pela internet, transformaram radicalmente os modos de produção científica e as dinâmicas do mercado de trabalho. O aumento exponencial na capacidade de coleta, armazenamento e processamento de dados — o chamado “dilúvio de dados” — criou um cenário em que a simples disponibilidade de informação não é mais suficiente: tornou-se necessário desenvolver métodos capazes de extrair, de forma eficiente e precisa, valor e conhecimento desses grandes volumes de dados (Levine, 2009; Finzer, 2013; Rautenberg; Carmo, 2019).

É nesse contexto que surge a Ciência de Dados (*Data Science*), consolidada como um quarto paradigma da investigação científica (Economist, 2017). Ao integrar conhecimentos de matemática, estatística, Inteligência Artificial (IA) e engenharia de computação, a Ciência de Dados oferece suporte metodológico para capturar, explorar e minerar dados massivos (*big data*¹), transformando fenômenos complexos em insights acionáveis. Essa abordagem multidisciplinar tem se expandido para diversas áreas, de biologia a ciências sociais, e tem gerado novos perfis profissionais capazes de navegar com fluidez entre teoria estatística e implementação computacional.

No campo da Economia, a aplicação de técnicas de *Machine Learning (ML)*² (aprendizado de máquina) tem promovido uma verdadeira revolução metodológica.

Ferramentas de ML permitem a construção de modelos preditivos avançados, que

¹ Refere-se a conjuntos de dados extremamente grandes e complexos, que não podem ser tratados com ferramentas tradicionais de processamento de dados, exigindo soluções avançadas para coleta, armazenamento e análise.

² Machine Learning (Aprendizado de Máquina) refere-se a um subconjunto da Inteligência Artificial que desenvolve algoritmos capazes de aprender padrões em dados e realizar previsões ou classificações automaticamente, sem ser explicitamente programado para cada tarefa.

aprendem automaticamente padrões em grandes bases de dados e realizam previsões com alta acurácia — indo além do escopo tradicional da econometria, cujo enfoque central é a inferência causal e a interpretação das relações entre variáveis (Monard; Baranauskas, 2003; Chollet; Allaire, 2017; Morettin; Singer, 2022). Ainda que os objetivos dessas duas abordagens sejam distintos — predição versus explicação —, a interseção entre elas tem se revelado extremamente produtiva, gerando pesquisas mais robustas, seguras e inovadoras (Athey; Luca, 2019; Lechner, 2023).

Diante desse panorama, é fundamental mapear a evolução das aplicações de ML na pesquisa econômica. A análise bibliométrica se apresenta como ferramenta ideal para essa tarefa, pois permite quantificar o progresso científico, identificar tendências emergentes e analisar redes de colaboração entre autores e instituições. Por meio de métricas como número de publicações, frequência de citações e estruturas de coautoria, é possível vislumbrar o alcance e a profundidade dos estudos que unem Ciência de Dados e Economia.

Este trabalho propõe-se a realizar uma revisão bibliométrica abrangente das publicações sobre ML em Economia, considerando o período de 1996 (primeiro registro encontrado) a 2024, nas bases *Scopus* e *Web of Science (WoS)*. Serão utilizadas ferramentas do *software R*³ e a interface *Biblioshiny*⁴ do pacote *bibliometrix*⁵, aplicando descritores como “machine learning” em combinação com termos econômicos (“economics”, “economy” etc.), para gerar um conjunto inicial trabalhos.

Além desta introdução, este estudo se divide em mais 4 seções. A próxima

³R Core Team (2022)

⁴É uma interface gráfica interativa do pacote Bibliometrix, que facilita o uso das ferramentas de análise bibliométrica sem necessidade de programação.

⁵É um pacote da linguagem R⁶ utilizado para realizar análises bibliométricas, como mapeamento de redes de autores, coautorias, cocitações e evolução de temas científicos Aria e Cuccurullo (2017).

apresenta uma breve contextualização sobre aplicações de Ciências de Dados em Economia, dando evidência a todo o debate atual existente acerca da utilização de modelos preditivos de ML e os resultados decorrentes. Já a terceira seção, evidencia os procedimentos metodológicos, incluindo a descrição dos dados e a estratégia empírica. A quarta seção apresenta os resultados da pesquisa, enquanto a quinta seção expõe as considerações finais.

2 MACHINE LEARNING EM ECONOMIA: UMA NOVA FRONTEIRA METODOLÓGICA

Esta seção aprofunda a discussão sobre a crescente simbiose entre *Machine Learning* (ML) e a Economia. Para além de uma simples ferramenta, o ML representa uma mudança paradigmática que redefine a atuação profissional do economista e expande as fronteiras da análise empírica. Conforme aponta Petrová (2022), a revolução tecnológica exige que a qualificação dos novos economistas seja ampliada para incorporar esses mecanismos, que são cruciais para a competitividade econômica. Nesse sentido, a literatura recente tem se dedicado a explorar como essa integração pode ser realizada de forma produtiva, preservando o rigor da análise econômica tradicional e, ao mesmo tempo, potencializando-a com o poder computacional do ML (Athey; Imbens, 2019).

2.1 A REVOLUÇÃO NA INFERÊNCIA CAUSAL E NA AVALIAÇÃO DE POLÍTICAS PÚBLICAS

O cerne de grande parte da pesquisa econômica aplicada reside na inferência causal — a difícil tarefa de isolar o efeito de uma variável sobre outra. A avaliação de políticas públicas, por exemplo, depende crucialmente da capacidade de estimar o que teria acontecido na ausência de uma intervenção. É neste campo que o ML tem provocado os avanços mais significativos. O estudo seminal de Lechner (2023), intitulado *Causal Machine Learning and its use for public policy*, destaca como os sistemas inteligentes estão elevando a qualidade das análises de impacto. A principal vantagem, segundo o autor, não é apenas gerar estimativas mais precisas, mas também

identificar a heterogeneidade dos efeitos de tratamento. Em outras palavras, os algoritmos de *Causal Machine Learning* permitem investigar detalhadamente para quais subgrupos da população uma política funciona melhor ou pior, possibilitando um uso mais eficiente dos recursos públicos e um aprimoramento contínuo das ações governamentais.

Na mesma linha, Athey e Imbens (2019) argumentam que uma das aplicações mais inovadoras do ML é justamente a estimação de Efeitos Médios de Tratamento (ATE), um parâmetro onipresente em análises micro e macroeconômicas. Os métodos de ML superam os econométricos tradicionais especialmente em cenários de alta dimensionalidade (muitas variáveis) ou quando as relações de tratamento são complexas e não-lineares. Indo além, os autores mostram o potencial do ML para a estimação de políticas ótimas. Em vez de apenas avaliar uma política existente, esses métodos permitem que os pesquisadores modelem a complexidade das interações entre variáveis para desenhar e sugerir intervenções potencialmente mais eficazes e precisas desde o início.

2.2 APLICAÇÕES EM MACROECONOMIA E ANÁLISE DE RISCO FINANCEIRO

Para além da inferência causal, o poder preditivo e de classificação do ML tem encontrado terreno fértil na macroeconomia e nas finanças, áreas que lidam com a previsão de agregados econômicos e a gestão de riscos sistêmicos. Um exemplo notável é o trabalho de Jesus e Besarria (2023), que aplicou a técnica não-supervisionada de *K-means* para criar novas métricas de risco de insolvência bancária no Brasil. A capacidade do algoritmo de processar dados não estruturados e identificar padrões (*clusters*) não-lineares nas características dos bancos permitiu uma classificação de

risco mais eficaz do que os métodos tradicionais.

A superioridade preditiva do ML em cenários de crise também é o foco de Casabianca et al. (2022). Os autores compararam um modelo econométrico padrão (logit) com um algoritmo de ML supervisionado (*AdaBoost*) para prever crises bancárias em uma amostra de países. Os resultados foram inequívocos: o modelo de ML apresentou um desempenho superior em todas as métricas relevantes de previsão fora da amostra ("AUROC"¹, sensibilidade e especificidade), demonstrando sua robustez para atuar como um sistema de alerta precoce.

A fronteira da pesquisa macroeconômica também explora técnicas mais complexas. Bitetto, Cerchiello e Mertzanis (2023), por exemplo, utilizaram a modelagem de elevação (*Uplift Modelling*)², uma área emergente do ML causal, para estimar o impacto individual de políticas fiscais expansionistas sobre o crescimento do Produto Interno Bruto (PIB) real. A abordagem permitiu identificar efeitos causais defasados e estimar o impacto granular das políticas, gerando resultados que, segundo os autores, são consistentes tanto com a teoria macroeconômica quanto com a realidade observada.

2.3 FRONTEIRAS, DESAFIOS E UMA VISÃO CRÍTICA

Apesar do entusiasmo, a adoção do ML não é isenta de desafios e críticas. No campo das finanças corporativas, Khan et al. (2023) utilizaram ML para prever a vulnerabilidade de empresas durante a pandemia de COVID-19. Embora tenham obtido

¹É uma métrica de desempenho usada para avaliar a capacidade de classificação de modelos preditivos, variando de 0,5 (aleatório) a 1 (classificação perfeita).

²Uplift Modelling é uma técnica de aprendizado de máquina usada para estimar o efeito incremental de uma ação (como uma política pública) sobre diferentes grupos de indivíduos.

previsões mais precisas, os próprios autores alertam para os riscos. A natureza de "caixa-preta" (*black box*³) de alguns algoritmos complexos pode dificultar a interpretação dos resultados, e as suposições subjacentes ao modelo podem gerar riscos para a validação e supervisão, como no caso da previsão de default de crédito. Essa visão crítica é fundamental, pois ressalta que o poder preditivo deve ser sempre balanceado com a interpretabilidade e a robustez teórica.

Em suma, o diálogo entre ML e Economia é vibrante e multifacetado. Os estudos levantados demonstram que os algoritmos de ML não são apenas ferramentas para previsão, mas mecanismos que potencializam a análise causal, a classificação de riscos e o desenho de políticas. Evidencia-se, portanto, que a conjuntura de avanços da Ciência de Dados está equipando os economistas com um arsenal metodológico mais poderoso, capaz de extrair insights mais profundos da crescente abundância de dados e, assim, aprimorar a tomada de decisões nos setores público e privado.

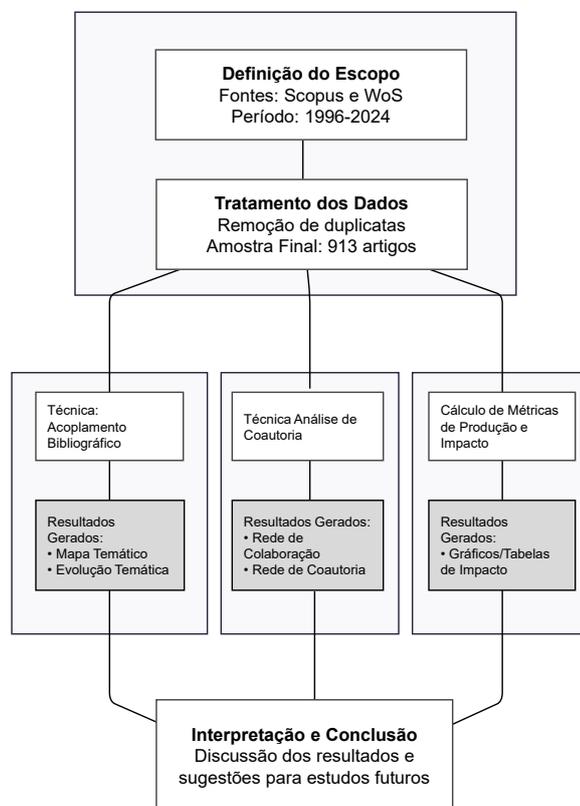
³Em ciência de dados, o termo *black box* refere-se a modelos cujos processos internos não são facilmente interpretáveis, dificultando a compreensão de como determinadas previsões são feitas.

3 PROCEDIMENTOS METODOLÓGICOS

O presente estudo adota como procedimento metodológico a análise bibliométrica, uma abordagem quantitativa robusta para mapear a estrutura intelectual e a evolução de um campo científico Broadus (1987). O objetivo é ir além de uma revisão de literatura tradicional, utilizando indicadores de atividade científica para identificar padrões, tendências, principais autores, periódicos de impacto e redes de colaboração na interseção entre ML e Economia Macias-Chapula (1998). Todo o procedimento, desde a coleta até a análise final dos dados, foi executado de forma a garantir a transparência e a reprodutibilidade da pesquisa.

A fim de proporcionar uma compreensão clara e sintética do fluxo de trabalho adotado nesta pesquisa, a Figura 1 ilustra as macroetapas sequenciais que nortearam o estudo. O processo, representado visualmente, parte do planejamento e da coleta de dados, avança para o tratamento e a consolidação da amostra, culmina na fase de análise bibliométrica e finaliza com a interpretação dos resultados e a elaboração das conclusões. Esta representação esquemática serve como um guia visual para a metodologia detalhada ao longo desta seção.

Figura 1 – Fluxograma metodológico



Fonte: Elaboração própria.

3.1 COLETA E DELIMITAÇÃO DA AMOSTRA

A primeira etapa consistiu na definição das fontes de dados. Foram selecionadas as bases de dados *Scopus Elsevier (2024)* e *WoS Clarivate (2024)*. A escolha conjunta das bases *Scopus* e *Web of Science* foi motivada por critérios de abrangência, padronização e compatibilidade técnica, especialmente com a linguagem R e o pacote *Bibliometrix*. Essas plataformas reúnem periódicos de alto impacto e possuem um rigoroso processo de indexação, o que garante a qualidade e a relevância dos documentos. Outro ponto fundamental foi a disponibilidade de metadados estruturados e homogêneos, permitindo análises mais precisas de acoplamento bibliográfico,

redes de coautoria, cocitação e mapas temáticos. A combinação dessas bases oferece, portanto, um panorama global consistente, ao mesmo tempo em que assegura a comparabilidade entre estudos e a replicabilidade dos resultados.

A coleta de dados foi realizada por meio de buscas avançadas, utilizando descritores específicos para capturar a literatura pertinente. As *strings* de busca¹ foram as seguintes:

- **Scopus:** (TITLE-ABS-KEY ("Machine Learning") AND TITLE-ABS-KEY ("economy" OR "Economics")) AND (LIMIT-TO (SUBJAREA , "ECON"))
- **Web of Science:** TS=("Machine Learning") AND TS=("Economy" OR "Economic*") AND WC=("Economics")

Esses parâmetros foram desenhados para garantir que os artigos recuperados tivessem alta relevância temática, pertencendo à área específica da Economia. A busca inicial retornou um total bruto de 1.666 documentos (913 da *Scopus* e 753 da *WoS*). Em seguida, as duas bases foram unificadas e processadas para a remoção de duplicatas. Após este tratamento, a amostra final consolidada para a análise resultou em 913 artigos únicos, cujas variáveis e escopo são detalhados no Quadro 1.

¹Strings de busca são comandos formados por uma combinação estratégica de palavras-chave, termos e operadores lógicos (como AND, OR, *) utilizados para executar uma pesquisa precisa e recuperar informações altamente relevantes de bases de dados ou motores de busca.

Quadro 1: Descrição das variáveis analisadas

Variáveis Analisadas	Nomenclatura	Descrição
Autor(es)	AU	Refere-se ao(s) autor(es) de um documento científico.
Afiliação	AF	Indica a(s) instituição(ões) à qual(is) o(s) autor(es) está(ão) associado(s).
Título	TI	O título do documento científico.
Ano de Publicação	PY	O ano em que o documento foi publicado.
Fonte	SO	A revista ou conferência onde o documento foi publicado.
Número Total de Citações	TC	O total de citações que o documento recebeu.
Resumo	AB	Um resumo do conteúdo do documento.
Descrição do Documento	DE	A descrição geral do documento, podendo incluir detalhes adicionais.
Tipo de Financiamento	FX	A categoria do financiamento recebido pelo estudo.
Localização	LA	A localização geográfica relacionada ao estudo ou aos autores.
Identificação do Journal	Jl	O identificador do periódico onde o documento foi publicado.
Autoria Universitária	AU_UN	Indica se o autor é vinculado a uma universidade.
Citações por Ano	TCpY	O número médio de citações recebidas por ano desde a publicação.
Número Total de Citações	NTC	O total de citações acumuladas do documento em todas as bases de dados.

Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

3.2 TRATAMENTO DOS DADOS E ANÁLISE ESTRUTURAL

O tratamento e a análise da amostra consolidada foram conduzidos no *software* R R Core Team (2022), com o auxílio do pacote *bibliometrix* e sua interface interativa *biblioshiny* Cobo, López-Cózar e Herrera-Viedma (2019), Aria e Cuccurullo

(2017). Para investigar a estrutura intelectual do campo, foram empregadas duas técnicas centrais de análise de redes: o Acoplamento bibliográfico e a análise de colaboração.

O Acoplamento bibliográfico, proposto originalmente por Kessler (1963), ocorre quando dois documentos (artigos A e B) citam um terceiro documento em comum (artigo C) em suas listas de referência. A força desse acoplamento é proporcional ao número de referências que compartilham, o que sugere uma alta correlação semântica entre eles. Essa análise é fundamental para identificar os "clusters" ou frentes de pesquisa dentro de um campo. A rede de acoplamento pode ser representada formalmente por uma matriz B , obtida pela seguinte formulação:

$$B = A \times A^T \quad (3.1)$$

onde A é a matriz de citação (documentos \times referências). Cada elemento b_{ij} da matriz simétrica B indica a força do acoplamento entre os documentos i e j . Para agrupar os documentos em clusters temáticos, o *bibliometrix* emprega algoritmos de clusterização, como o "*walktrap*"², sobre essa matriz.

Para mapear a estrutura social da pesquisa, foi analisada a rede de colaboração entre autores. Esta análise revela os grupos de pesquisadores que trabalham em conjunto e as pontes entre diferentes escolas de pensamento. A rede de coautoria é obtida a partir da seguinte formulação geral:

$$AC = A_{ca}^T \times A_{ca} \quad (3.2)$$

²O algoritmo *walktrap* é um método de detecção de comunidades em redes complexas baseado em caminhadas aleatórias, utilizado para identificar grupos coesos de elementos interconectados.

onde A_{ca} é uma matriz bipartida que relaciona os Manuscritos \times Autores. O elemento c_{ij} da matriz resultante AC indica a força da colaboração entre o autor i e o autor j , medida pelo número de publicações em coautoria.

A partir dessas técnicas, buscou-se extrair e analisar os principais indicadores bibliométricos, como a evolução da produção anual, os autores e instituições mais prolíficos, os periódicos de maior impacto e, principalmente, as tendências temáticas emergentes. A interpretação desses resultados permite construir um panorama fiel do avanço do ML no campo da Economia.

4 RESULTADOS

A análise bibliométrica sobre a aplicação de ML na Economia revela um campo de pesquisa em estado de efervescência, marcado por um crescimento explosivo e um caráter colaborativo global. O painel geral, apresentado na Figura 2, sintetiza os principais indicadores deste cenário. A publicação de 913 documentos no período de 1996 a 2024, com uma significativa taxa de crescimento anual de 21,42%, sinaliza uma aceleração acentuada no interesse pela área. Este dinamismo é corroborado pela baixa idade média dos documentos, de apenas 2,07 anos, indicando que a vasta maioria da produção científica é muito recente. O impacto dessa produção na comunidade acadêmica é igualmente notável, com uma média de 14,48 citações por documento, o que reflete a alta relevância e a rápida absorção dos avanços neste campo interdisciplinar.

Figura 2 – Painel Geral



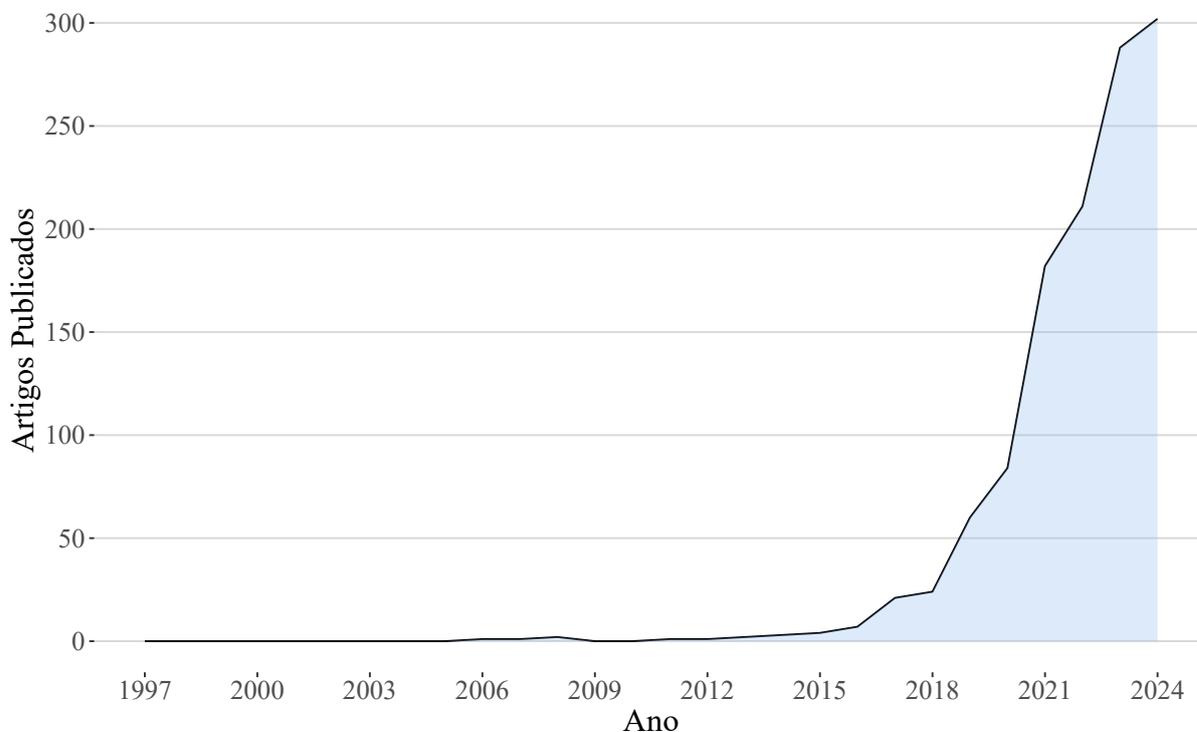
Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

O perfil da comunidade científica que impulsiona este avanço é predominantemente colaborativo. A média de 3,05 coautores por documento demonstra que a pesquisa na área é, em sua essência, um esforço de equipe, característica comum

em áreas que exigem expertises diversas, como Economia e Ciência da Computação. Embora existam 125 trabalhos de autoria única, a tendência é a cooperação, que transcende fronteiras geográficas: 14,79% dos artigos envolvem coautorias internacionais, sublinhando a natureza globalizada da investigação. A profundidade e a amplitude temática deste esforço colaborativo são evidenciadas pelo robusto alicerce teórico, com um total de 48.557 referências citadas, e pela diversidade de tópicos, indicada pelas 2.928 palavras-chave distintas fornecidas pelos autores.

A análise da produção científica ao longo do tempo, detalhada no Gráfico 1, materializa e explica a relevante taxa de crescimento de 21,42% ao ano, identificada previamente no painel geral. O gráfico revela uma trajetória marcada por três fases distintas. A primeira, uma fase latente que se estende de 1996 até aproximadamente 2016, caracteriza-se por uma atividade acadêmica esparsa, com poucas publicações anuais, indicando que o tema era um nicho restrito de pesquisa.

A partir de 2017, observa-se um claro ponto de inflexão, onde a trajetória de crescimento muda drasticamente. Este marco dá início à terceira fase, de crescimento exponencial, que se intensifica a cada ano até 2024. O impacto visual dessa aceleração é notável: a grande maioria de toda a produção científica do campo está concentrada nos últimos seis anos. Esta concentração de publicações em um passado tão recente explica diretamente outro indicador-chave do painel geral: a baixíssima idade média dos documentos, de apenas 2,07 anos. Portanto, o gráfico não apenas ilustra um aumento no volume de publicações, mas também confirma que a intersecção entre ML e Economia é um campo de pesquisa que, após um longo período de maturação, crescimento expressivo e teve intensa atividade na última década.

Gráfico 1: Evolução das publicações ao longo do tempo

Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

A análise dos periódicos científicos revela os principais canais de disseminação do conhecimento e, mais importante, aponta para as áreas da Economia onde a aplicação de ML tem se consolidado com maior vigor. A Tabela 1 classifica os 10 periódicos de maior impacto de acordo com o Índice H¹, uma métrica que pondera tanto a produtividade quanto o número de citações de um periódico ou autor. A análise dos títulos permite identificar três grandes eixos temáticos que dominam o cenário: Previsão, Energia e Recursos Naturais, e Modelagem Computacional.

O eixo de Previsão é liderado pelo *International Journal of Forecasting* (1^o, H-

¹O Índice H avalia a produtividade e o impacto de um pesquisador ou periódico com base em suas publicações, buscando equilibrar a quantidade de publicações e a quantidade de citações que elas recebem.

Index 13) e pelo *Journal of Forecasting* (6º, *H-Index* 7). A proeminência destes periódicos é natural, visto que as técnicas de ML são inerentemente preditivas e oferecem avanços significativos sobre os modelos econométricos clássicos de séries temporais. O segundo grande polo temático é o de Energia e Recursos Naturais, representado por uma quantidade expressiva de periódicos na lista, como *Resources Policy* (2º, *H-Index* 11), *Energy Economics* (3º, *H-Index* 10), e *Resources, Conservation and Recycling* (5º, *H-Index* 8). Esta concentração evidencia o intenso uso de ML para modelar mercados de commodities, otimizar o consumo energético e analisar políticas ambientais, problemas complexos e ricos em dados.

Tabela 1 – Principais Periódicos que publicam artigos de *Machine Learning* Aplicado à Economia

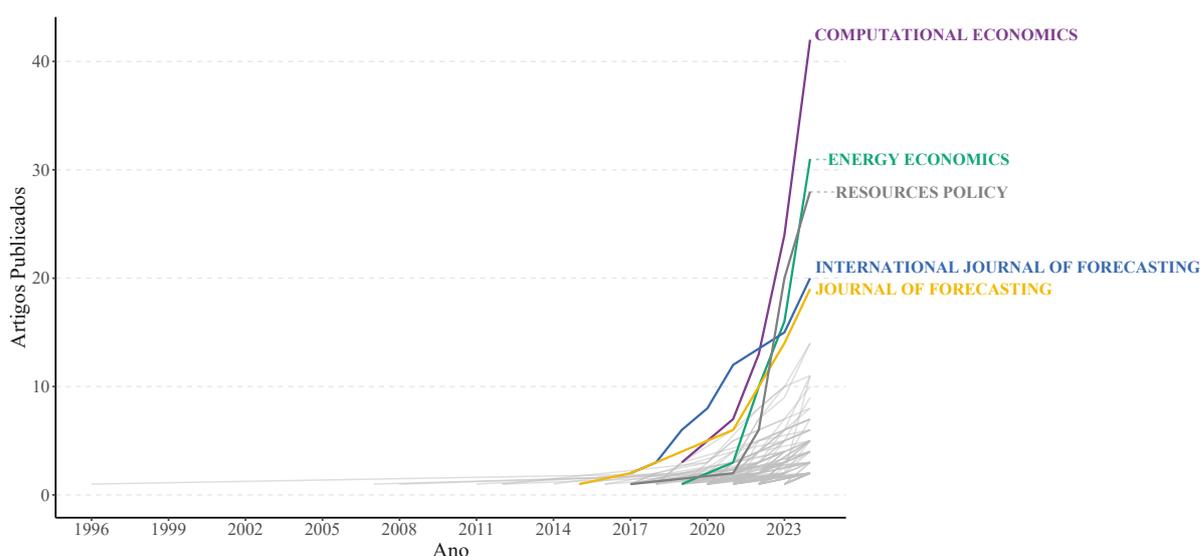
#	Periódico	H Index	Citações	Artigos
1	International Journal of Forecasting	13	613	20
2	Resources Policy	11	559	28
3	Energy Economics	10	523	31
4	Computational Economics	9	297	42
5	Resources, Conservation and Recycling	8	246	14
6	Journal of Forecasting	7	157	19
7	Applied Economics	6	134	14
8	Economic Modelling	6	199	10
9	Frontiers in Energy Research	5	71	10
10	Regional Science and Urban Economics	5	75	7

Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

Uma observação crucial emerge ao comparar o ranking de impacto com o de volume. O periódico *Computational Economics*, embora em 4º lugar no *H-Index*, é o líder absoluto em quantidade, com 42 artigos publicados. Esta aparente discrepância é perfeitamente elucidada pela análise dinâmica do Gráfico 2. O gráfico mostra que, enquanto os periódicos de previsão foram os pioneiros na adoção do tema, o *Computati-*

onal Economics apresentou um crescimento explosivo e mais recente, consolidando-se rapidamente como o principal veículo para um grande volume de pesquisas na área.

Gráfico 2: Publicação dos periódicos ao longo do tempo



Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

O Gráfico 2 também detalha a trajetória dos líderes temáticos. Após um período de baixa atividade até 2015, vemos o *Journal of Forecasting* iniciar o movimento, seguido de perto pelo *International Journal of Forecasting*. Contudo, a partir de 2019, o crescimento dos periódicos de energia e, sobretudo, do *Computational Economics*, redefine o panorama, demonstrando a diversificação e o amadurecimento do campo. Em síntese, a análise das fontes não apenas identifica os periódicos mais relevantes, mas mapeia a própria estrutura da pesquisa, revelando como diferentes subáreas da Economia estão, cada uma a seu ritmo, absorvendo e contribuindo para a revolução do ML.

As instituições de pesquisa e ensino superior são o epicentro da inovação, e

a análise de suas contribuições permite mapear onde se concentram os recursos e talentos que impulsionam a fronteira do conhecimento em ML aplicado à Economia. A Tabela 2 apresenta as dez organizações mais influentes, classificadas pelo total de citações de seus trabalhos.

Tabela 2 – Principais Instituições em Pesquisa de *Machine Learning* na Economia

Posição	Instituição	Citações	Rank	Setor
1	National Bureau of Economic Research	1250	9	Sem fins lucrativos
2	University of Chicago	1162	8	Universidades
3	Harvard University	1084	8	Universidades
4	Yale University	920	7	Universidades
5	University of California System	861	12	N/A
6	University of California Berkeley	710	5	Universidades
7	Google Incorporated	637	1	Empresas
8	Center for Economic and Policy Research	546	5	N/A
9	Cornell University	363	7	Universidades
10	Stanford University	331	3	Universidades

Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

Uma análise aprofundada dos dados revela duas conclusões centrais. A primeira é uma notável concentração geográfica: todas as dez instituições que lideram em impacto global são sediadas nos Estados Unidos. Este fato sugere uma liderança hegemônica do país no desenvolvimento e na aplicação dessas tecnologias no campo econômico. A segunda conclusão é a formação de um ecossistema de pesquisa composto por um tripé de atores distintos. Universidades de elite, como a University of Chicago (1162 citações) e a Harvard University (1084 citações), formam a base acadêmica tradicional. A elas se somam institutos de pesquisa econômica altamente especializados e sem fins lucrativos, notavelmente o *National Bureau of Economic*

Research (NBER), que ocupa a primeira posição com 1250 citações.

O terceiro pilar deste ecossistema é o setor corporativo, representado de forma proeminente pelo *Google Incorporated*. A presença da empresa na lista, com 637 citações, evidencia a crescente intersecção entre a pesquisa acadêmica e a inovação industrial, onde os avanços em ML desenvolvidos no setor de tecnologia são aplicados para resolver problemas econômicos complexos.

Adicionalmente, os dados permitem uma análise sobre o perfil de impacto de cada instituição. A coluna "Rank", que presumivelmente indica o número de artigos publicados, mostra diferentes estratégias de publicação. Instituições como o NBER e o University of California System alcançam seu alto impacto através de um volume maior de publicações (9 e 12 artigos, respectivamente). Em contrapartida, o Google demonstra um modelo de altíssimo impacto concentrado, obtendo suas 637 citações a partir de um único artigo. Este padrão destaca que a influência no campo pode ser alcançada tanto por uma produção científica consistente quanto por contribuições singulares de grande repercussão.

Explorar os autores mais influentes no campo de ML aplicado à Economia é essencial para identificar os pesquisadores que lideram o avanço do conhecimento e da inovação. Esta análise revela não apenas quem são os líderes, mas também suas trajetórias de produção e as Redes de colaboração que sustentam o avanço científico. Para isso, foi realizada uma análise multifacetada, combinando um ranking de impacto, um mapa da produção temporal e uma rede de coautorias.

A Tabela 3 apresenta os pesquisadores de maior impacto, classificados pelo Índice H. Observa-se que Gupta R lidera com um Índice H de 5, acumulando 122 citações em 6 artigos desde 2015. Em seguida, um grupo de pesquisadores, como Kim

S, Li J e Wang Y, compartilha um Índice H de 4. Um ponto fundamental que emerge desta tabela é o nível relativamente baixo dos Índices H. Em campos consolidados, os principais autores frequentemente possuem índices muito mais elevados. Este dado é um forte indicativo de que estamos analisando um campo de pesquisa muito jovem, onde mesmo os pesquisadores mais impactantes ainda estão em fase de acumulação de citações ao longo do tempo.

Tabela 3 – Autores mais impactantes em artigos que versam sobre *Machine Learning* aplicado à Economia

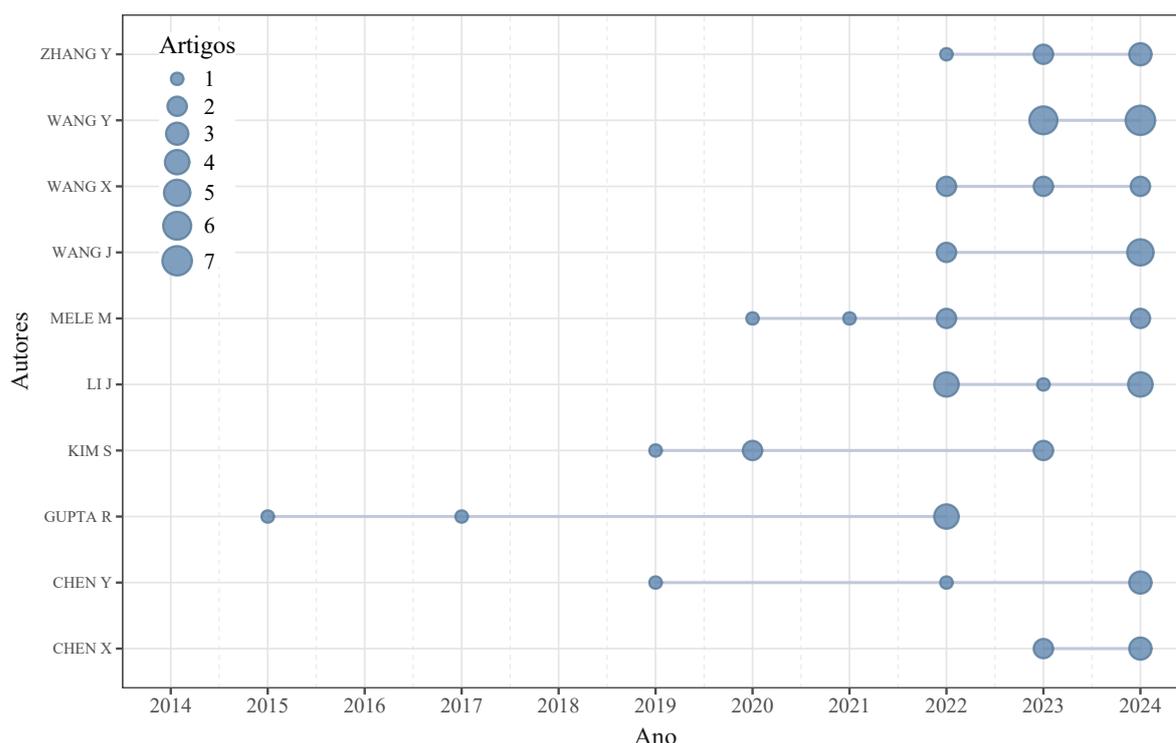
Classificação	Autor	H Index	Citações	Artigos	Ano Inicial
1	Gupta R	5	122	6	2015
2	Kim S	4	62	5	2019
3	Li J	4	56	9	2022
4	Li X	4	35	5	2021
5	Pierdzioch C	4	123	5	2017
6	Wang Y	4	118	13	2023
7	Zhang C	4	76	5	2021
8	Chen X	3	26	5	2023
9	Chen Y	3	20	5	2019
10	Gogas P	3	88	4	2015

Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

Se a Tabela 3 nos oferece uma fotografia estática do impacto, o Gráfico 3 nos proporciona um filme, detalhando a dinâmica da produção científica de cada autor. Este tipo de gráfico, conhecido como gráfico de dispersão temporal ou gráfico de bolhas, mapeia a produção no eixo do tempo (eixo x), onde cada ponto representa uma publicação e o tamanho da bolha indica o volume de artigos publicados naquele ano específico. A visualização revela diferentes perfis de pesquisadores. De um lado, temos os "pioneiros", como Gupta R e Gogas P, cujas publicações se iniciam em 2015 e se distribuem de forma consistente ao longo dos anos. De outro lado, emerge uma

"nova onda" de pesquisadores, como Wang Y, Li J, e Chen X, que iniciaram suas publicações muito mais recentemente (a partir de 2021-2023) mas com uma intensidade e volume notavelmente altos, refletindo a aceleração do campo discutida anteriormente.

Gráfico 3: Principais autores ao longo do tempo

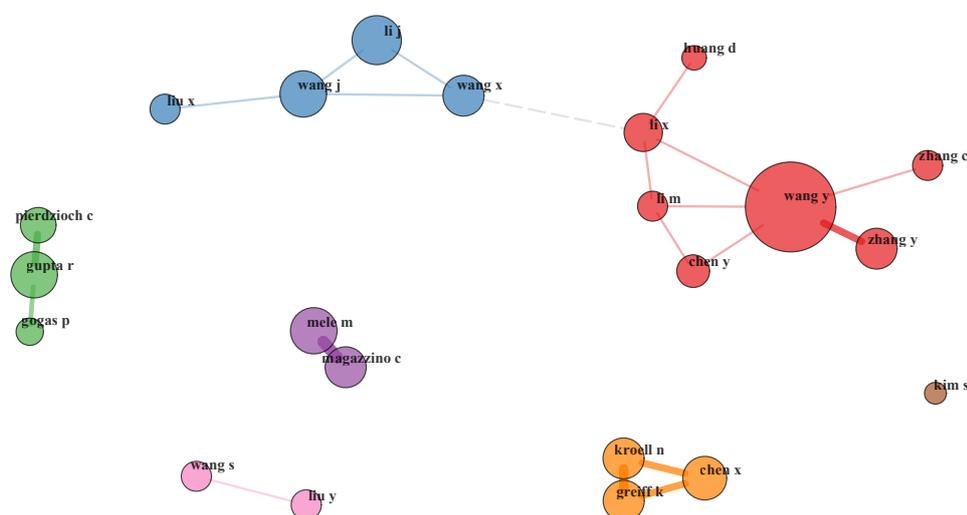


Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

Além do impacto individual e da trajetória temporal, é crucial entender como esses pesquisadores se conectam para gerar conhecimento. A Figura 3 elucidada essa estrutura social por meio de um mapa de colaboração. Neste tipo de visualização de rede, cada autor é um nó (círculo), e uma linha conectando dois autores significa que eles publicaram pelo menos um trabalho em coautoria. Os autores são agrupados por cor em "*clusters*", que representam grupos de pesquisa que colaboram intensamente entre si.

A principal revelação desta rede não é apenas quem colabora, mas a sua estrutura geral: ela é altamente fragmentada. Em vez de uma grande comunidade interconectada, o campo se organiza em múltiplas "ilhas" ou "*clusters*" de pesquisa, com pouca ou nenhuma colaboração entre eles. Podemos identificar claramente os principais grupos: um grande *cluster* em vermelho (liderado por Wang Y, Zhang Y e Chen Y), um em azul (em torno de Li J, Wang J e Wang X), e um em verde (conectando os pioneiros Pierdzioch C, Gupta R e Gogas P). A existência desses grupos isolados sugere que a pesquisa está avançando em frentes paralelas, possivelmente com especializações temáticas distintas dentro de cada *cluster*. A forte presença de autores com afiliações em países como China e Estados Unidos dentro desses grupos reflete onde os principais centros de pesquisa e inovação nesta área estão localizados.

Figura 3 – Rede de colaboração internacional entre autores em produções científicas



Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

Em síntese, a análise combinada das três visualizações pinta um retrato detalhado de um campo emergente: ele é liderado por pesquisadores com perfis distintos

(pioneiros de longa data e uma nova onda de alta intensidade), cujo impacto ainda está em processo de maturação. Crucialmente, esses pesquisadores operam em Redes de colaboração coesas, porém isoladas umas das outras, indicando uma fase de desenvolvimento científico em que diferentes escolas de pensamento ou especialidades estão evoluindo de forma independente antes de, possivelmente, convergirem no futuro.

A análise dos artigos mais citados permite identificar os trabalhos seminais que servem como alicerce para a pesquisa em ML aplicado à Economia. Estes documentos não apenas introduzem conceitos e métodos, mas definem a agenda de pesquisa, inspirando novas investigações. A Tabela 4 apresenta os dez trabalhos mais influentes, e uma análise de seus conteúdos revela que eles se organizam em torno de duas correntes principais: um grupo de artigos *metodológicos* que constroem a ponte entre ML e econometria, e um grupo de artigos de *aplicação*, predominantemente focados em finanças e outros domínios específicos.

Tabela 4 – Papers mais citados em artigos de economia que utilizam ML

Artigo	Identificador de Objeto Digital (DOI)	Citações	Citações anual
MULLAINATHAN S, 2017, J ECON PERSPECT	< https://doi.org/10.1257/jep.31.2.87 >	729	91.13
VARIAN HR, 2014, J ECON PERSPECT	< https://doi.org/10.1257/jep.28.2.3 >	637	57.91
GU SH, 2020, REV FINANC STUD	< https://doi.org/10.1093/rfs/hhaa009 >	608	121.60
ATHEY S, 2019, ANNU REV ECON	https://doi.org/10.1146	349	58.17
KLEINBERG J, 2018, Q J ECON	< https://doi.org/10.1093/qje/qjx032 >	306	43.71
MANELA A, 2017, J FINANC ECON	< https://doi.org/10.1016/j.jfineco.2016.01.032 >	261	32.63
GHODDUSI H, 2019, ENERGY ECON	< https://doi.org/10.1016/j.eneco.2019.05.006 >	256	42.67
NIKOU M, 2019, INTELL SYST ACCOUNT FINANCE MANAG	< https://doi.org/10.1002/isaf.1459 >	181	30.17
KOZAK S, 2020, J FINANC ECON	< https://doi.org/10.1016/j.jfineco.2019.06.008 >	173	34.60
CARMONA P, 2019, INT REV ECON FINANC	< https://doi.org/10.1016/j.iref.2018.03.008 >	163	27.17

Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

A primeira e mais influente corrente intelectual é a que busca estabelecer os fundamentos metodológicos para a integração do aprendizado de máquina nas ciências econômicas. O artigo mais citado da área, de Mullainathan e Spiess (2017) (729 citações), e o trabalho seminal de Varian (2014) (637 citações) são pilares desta discussão. Ambos argumentam que, diante da abundância de dados, o ML oferece ferramentas preditivas poderosas, mas alertam para a necessidade de uma aplicação cautelosa, distinguindo claramente as tarefas de predição das de inferência causal. Nesta mesma linha, o trabalho de Athey e Luca (2019) (349 citações), em linha com sua posição já identificada como uma das autoras de maior impacto, aprofunda a relação entre ML e econometria com foco em inferência causal e estimação de políticas.

Complementarmente, Kleinberg et al. (2018) (306 citações) exploram como os algoritmos preditivos podem aprimorar a tomada de decisão em políticas públicas, utilizando o sistema de fiança criminal como um estudo de caso sobre ganhos de bem-estar.

A segunda grande corrente de pesquisa dedica-se à aplicação destas técnicas em finanças, que se revela o campo mais fértil para a utilização de ML. O trabalho de Gu, Kelly e Xiu (2020) é emblemático deste movimento. Com 608 citações totais e uma taxa significativa de 121 citações anuais, ele é o estudo de maior impacto recente da lista, demonstrando a imensa relevância atual da aplicação de árvores de decisão e redes neurais para a previsão de prêmios de risco de ativos. Outros trabalhos de alto impacto seguem esta vertente, como o de Kozak, Nagel e Santosh (2020), que desenvolve um fator de desconto estocástico para prever retornos, o de Manela e Moreira (2017), que utiliza processamento de texto para medir a incerteza do mercado, e o de Carmona, Climent e Momparler (2019), que aplica algoritmos de *gradient boosting* para prever falências bancárias.

Finalmente, um terceiro grupo de artigos ilustra a aplicação de ML em domínios específicos e com dados complexos. O trabalho de Ghodduzi, Creamer e Rafizadeh (2019), por exemplo, revisa o uso de técnicas como SVM e redes neurais para previsão de preços de energia. É notável que este artigo tenha sido publicado no periódico *Energy Economics*, que a análise anterior já havia identificado como um dos principais polos de publicação do tema. Similarmente, Nikou, Mansourfar e Bagherzadeh (2019) focam na complexidade de séries temporais financeiras, aplicando aprendizado profundo para a previsão de preços de ações.

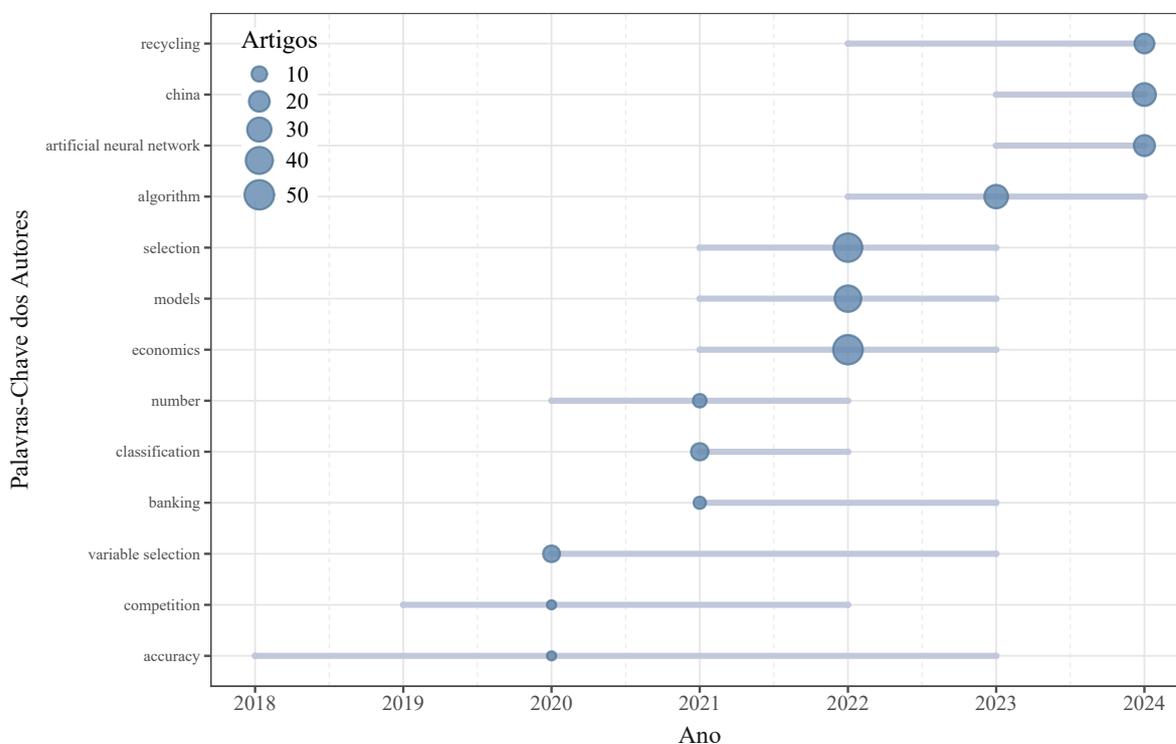
Em suma, a análise dos artigos mais influentes desenha um mapa claro do desenvolvimento do campo: ele foi fundado sobre uma base de trabalhos metodológi-

cos que o legitimaram academicamente, e seu crescimento tem sido impulsionado por aplicações de alto impacto, principalmente no setor financeiro e em outros domínios ricos em dados complexos.

A etapa seguinte desta análise bibliométrica mapeia o panorama intelectual do campo, identificando os tópicos de pesquisa que estão em ascensão e como eles se organizam em frentes de pesquisa coerentes. Para isso, foram utilizadas duas visualizações complementares: uma análise da evolução temporal das palavras-chave e um mapa temático que revela a estrutura de suas inter-relações. O estudo desses temas aprofunda o entendimento sobre as direções atuais e futuras da aplicação de ML na Economia.

O Gráfico 4 apresenta a evolução das palavras-chave mais proeminentes ao longo do tempo. Este tipo de visualização, um gráfico de dispersão temporal com bolhas, permite observar quais conceitos ganharam tração e em que momento. O tamanho da bolha representa a frequência do termo em um determinado ano. A análise revela uma maturação temática: os tópicos que surgem por volta de 2020, como *"accuracy"* (precisão), *"competition"* (competição) e *"variable selection"* (seleção de variáveis), representam preocupações fundamentais sobre a qualidade e a validação dos modelos. Posteriormente, a partir de 2021 e 2022, termos como *"banking"* (bancos) e *"classification"* (classificação) indicam uma aplicação mais direcionada. Mais recentemente, em 2023 e 2024, a ascensão de termos altamente específicos como *"artificial neural network"* (redes neurais artificiais), *"recycling"* (reciclagem) e *"china"* demonstra a consolidação de métodos avançados e o foco em contextos econômicos e ambientais particulares.

Gráfico 4: Principais tópicos emergentes



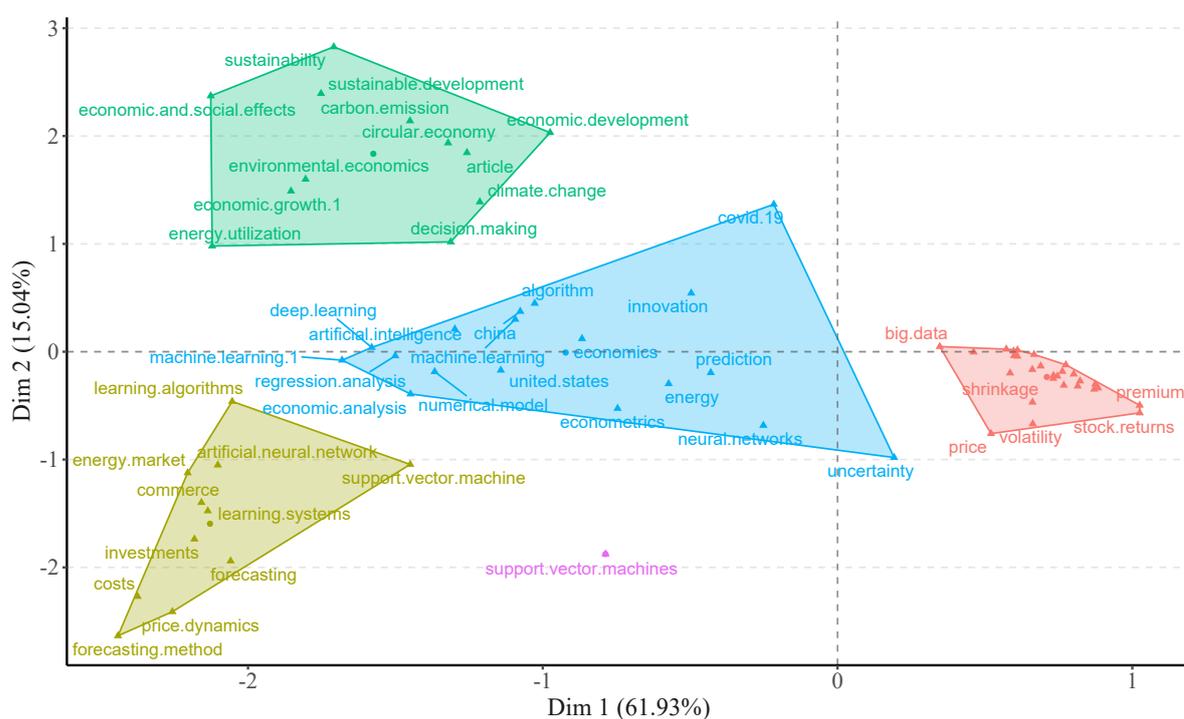
Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

Enquanto a análise temporal mostra quando os temas se tornaram relevantes, o mapa temático do Gráfico 5 revela como eles se estruturam e se conectam. Este tipo de diagrama posiciona os temas em um plano bidimensional com base em sua centralidade (importância e conexão com outros temas, no eixo horizontal) e densidade (desenvolvimento interno do tema, no eixo vertical). A análise dos agrupamentos (*clusters*) permite, assim, identificar as principais frentes de pesquisa do campo.

No coração do mapa, encontra-se o *cluster* azul, que funciona como o núcleo metodológico e conceitual da área. Ele contém os termos fundamentais como "ML", "deep learning", "econometrics", "prediction" e "algorithm". Sua posição central indica que todos os outros temas, de alguma forma, se conectam a este núcleo. À direita, o

cluster vermelho representa um dos principais temas motores da pesquisa: a fronteira da Finanças Quantitativa. Com termos como "stock returns", "volatility", "price" e "big data", este grupo é ao mesmo tempo central e bem desenvolvido, impulsionando muitas das aplicações práticas.

Gráfico 5: Mapa temático



Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

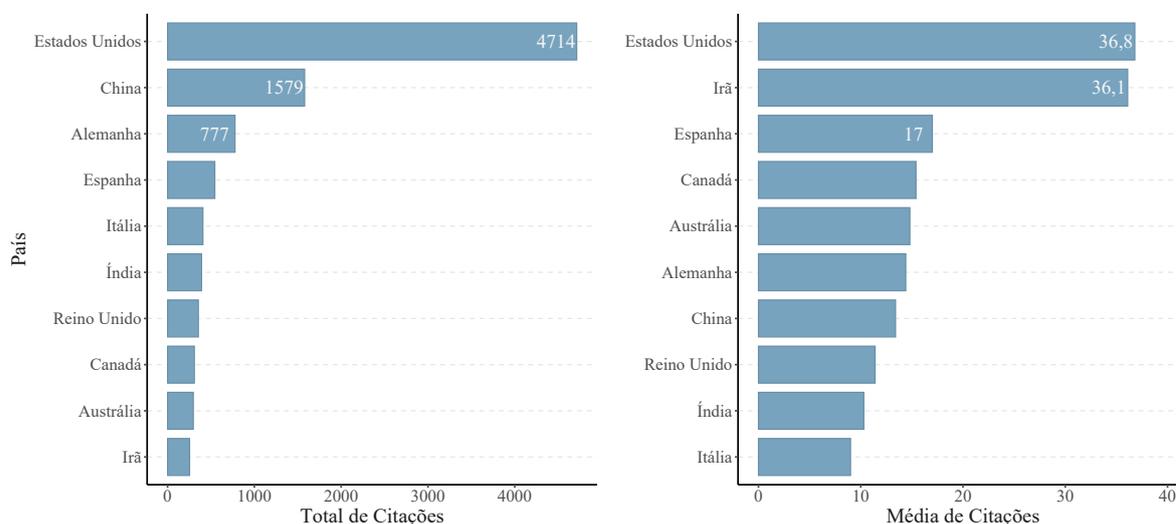
No canto superior esquerdo, o *cluster* verde emerge como um tema de nicho, altamente especializado: a Economia da Sustentabilidade. Termos como "sustainability", "circular economy", "carbon emission" e "climate change" formam um campo de pesquisa internamente coeso, mas, no momento, mais periférico em relação ao núcleo principal. A ascensão recente da palavra-chave "recycling" no Gráfico 4 sugere que este nicho está em plena expansão. Por fim, o *cluster* amarelo (canto inferior

esquerdo) agrupa temas básicos e transversais, como "forecasting", "investments" e "learning systems", que funcionam como ferramentas e conceitos aplicados nos outros *clusters*.

Dessa forma, a análise temática oferece um panorama da estrutura intelectual do campo, mostrando que a pesquisa em ML e Economia se organiza em torno de um núcleo metodológico que alimenta, principalmente, duas grandes frentes: uma área motora e já consolidada em finanças quantitativas e uma área de nicho, especializada e em forte crescimento, voltada para a sustentabilidade. Essa organização temática se conecta de forma consistente com os achados anteriores; a proeminência do *cluster* de finanças, por exemplo, espelha a forte presença de artigos sobre o tema entre os mais citados, enquanto o *cluster* de sustentabilidade se alinha à identificação de periódicos e palavras-chave emergentes nessa área.

A análise da distribuição geográfica da produção científica é fundamental para compreender onde se concentram os principais polos de pesquisa e como as nações interagem para fazer avançar o conhecimento. Esta seção explora a geografia da pesquisa em ML e Economia sob duas óticas complementares: o impacto das publicações de cada país e a estrutura da rede de colaboração internacional.

O Gráfico 6 oferece um panorama do impacto científico por meio de duas métricas distintas. O gráfico à esquerda, de total de citações, estabelece de forma clara a hegemonia dos Estados Unidos como a principal potência na área, com 4714 citações, um volume que supera em muito o da China (1579), segunda colocada, seguida por potências científicas europeias como Alemanha e Espanha. Este dado corrobora a análise anterior sobre as instituições, que já apontava para uma forte concentração de centros de pesquisa de elite nos EUA.

Gráfico 6: Países mais citados

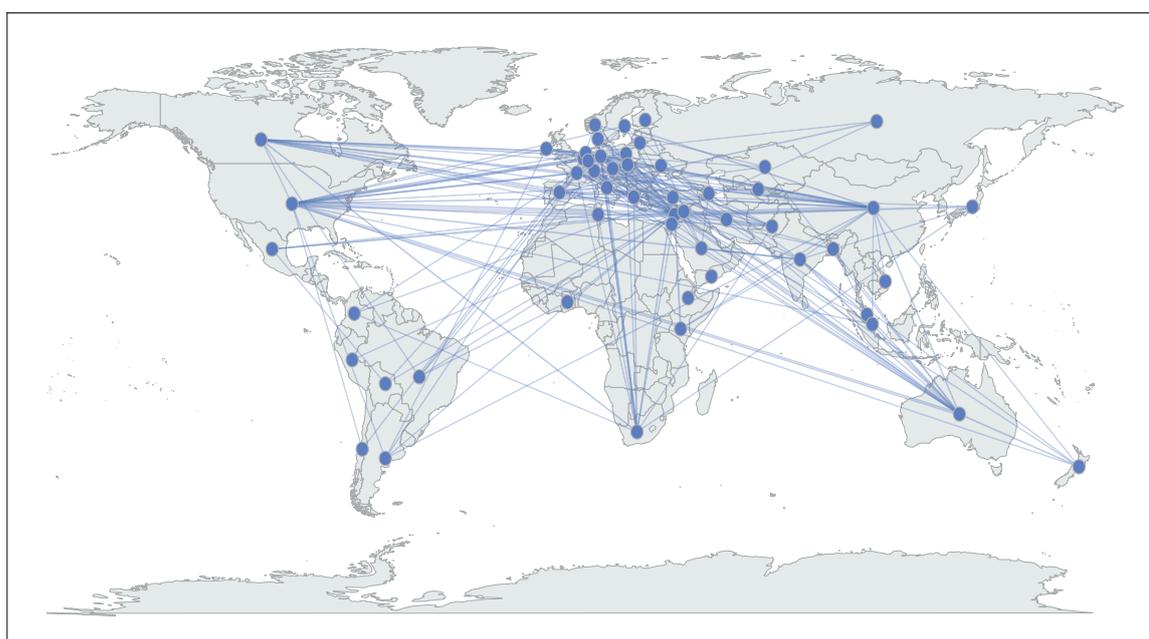
Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

Contudo, uma perspectiva distinta e reveladora emerge da análise da média de citações por artigo, no gráfico à direita. Embora os Estados Unidos mantenham a liderança com uma alta média de 36,8 citações, o dado mais surpreendente é a aparição do Irã em segundo lugar, com uma média quase idêntica de 36,1. Este resultado sugere que, apesar de o Irã possuir um volume de produção total menor neste campo específico, os poucos trabalhos provenientes do país que integram o corpus da pesquisa possuem um impacto significativamente alto, indicando a existência de um nicho de pesquisa de altíssima qualidade e relevância internacional. Esta dicotomia entre volume total e impacto médio ilustra que a influência científica global não é monolítica, havendo espaço para contribuições de alto impacto de nações fora do eixo tradicional.

Para além do impacto de cada país, a Figura 4 mapeia a dinâmica das colaborações que estruturam a pesquisa global. Nesta visualização de rede, cada ponto representa um país e as linhas simbolizam coautorias entre pesquisadores de dife-

rentes nações. A densidade e o alcance dessas linhas são cruciais para entender a interconexão da ciência, pois redes robustas aceleram a inovação. O mapa revela um padrão claro: um "triângulo" de intensa colaboração conecta a América do Norte, a Europa Ocidental e a Ásia Oriental.

Figura 4 – Rede de colaboração internacional entre países em produções científicas



Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

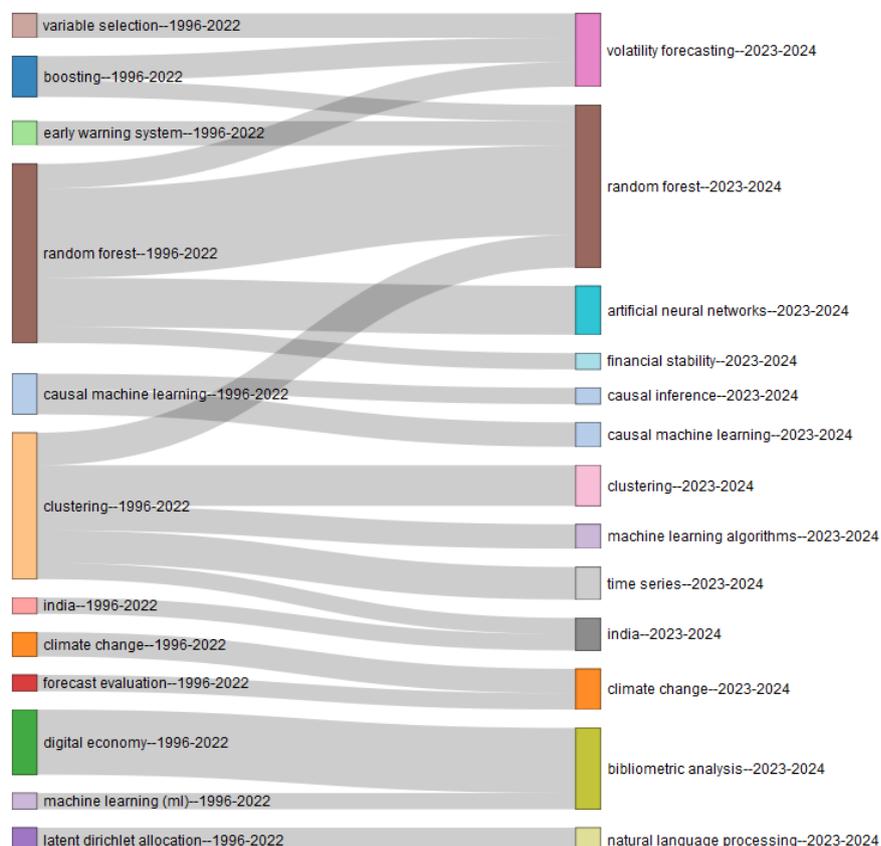
Dentro dessa estrutura, a Europa se destaca como um *cluster* de altíssima densidade interna, com inúmeras conexões entre seus países, funcionando como um bloco de pesquisa integrado. Os Estados Unidos, por sua vez, atuam como um "hub" global, sendo o país com as conexões mais diversificadas e de longo alcance, conectando-se de forma proeminente com todas as outras regiões do mundo. Em contrapartida, a menor densidade de conexões envolvendo países da África e da América Latina evidencia uma marcante desigualdade regional, refletindo uma menor partici-

pação relativa dessas regiões nas redes globais de produção de conhecimento sobre o tema. Assim, o mapa de colaboração demonstra tanto a forte interdependência dos principais polos científicos quanto as disparidades que ainda marcam o cenário científico internacional.

Para compreender a trajetória intelectual do campo, é essencial analisar não apenas os temas presentes, mas como eles evoluem, ganham ou perdem proeminência ao longo do tempo. A Figura 5 oferece um panorama desta dinâmica por meio de um diagrama de Sankey. Este tipo de visualização é particularmente poderoso para ilustrar fluxos e transições. No diagrama, as barras verticais representam dois períodos distintos: um período de fundação e desenvolvimento (1996-2022) à esquerda, e um período de maturação e sofisticação recente (2023-2024) à direita. A espessura de cada faixa temática indica sua relevância (frequência nos artigos), e as "correntes" que fluem da esquerda para a direita mostram como os focos de pesquisa se transformaram, persistiram ou deram origem a novos interesses.

No período inicial (1996-2022), a análise da coluna esquerda do diagrama revela uma concentração de esforços no estabelecimento da "caixa de ferramentas" do ML para a Economia. Temas como *variable selection*, *boosting*, *random forest* e *clustering* são proeminentes, indicando um foco intenso nos algoritmos fundamentais e nos desafios metodológicos para garantir a robustez das análises. Simultaneamente, aplicações iniciais já apontavam para áreas de grande potencial, como *early warning system* (sistemas de alerta precoce) para prever crises e o estudo da *digital economy*. A presença de temas como *climate change* e *india* desde essa fase inicial também mostra que certas questões ambientais e contextos geográficos específicos já eram relevantes.

Figura 5 – Evolução temática



Fonte: Elaboração própria a partir de Elsevier (2024) e Clarivate (2024).

A verdadeira riqueza da análise, contudo, reside nos fluxos que conectam os dois períodos. Observa-se uma clara trajetória de maturação, onde técnicas basilares evoluem para aplicações mais especializadas e sofisticadas. Por exemplo, o conhecimento consolidado em *boosting* e *variable selection* agora é canalizado diretamente para um problema mais complexo e de alto impacto: *volatility forecasting* (previsão de volatilidade). De forma similar, o conceito técnico de *early warning system* amadurece e se alarga para o objetivo macroeconômico de *financial stability* (estabilidade financeira). Outra transição notável é a do método *latent dirichlet allocation*² (uma téc-

²*Latent Dirichlet Allocation (LDA)* é um método estatístico para descobrir automaticamente tópicos laten-

nica específica de modelagem de tópicos) para o campo mais amplo e moderno de *natural language processing* (Processamento de Linguagem Natural), sinalizando a crescente importância de analisar dados textuais, como notícias e relatórios, para a análise econômica.

Ao mesmo tempo, o diagrama revela a persistência e a diversificação de temas centrais. O método *random forest*, por exemplo, continua sendo um tópico relevante por si só no período recente, mas também serve de base para o avanço em *artificial neural networks* (redes neurais artificiais), mostrando uma expansão do arsenal técnico dos pesquisadores. O tema *causal ML*, um dos mais robustos no período inicial, não só continua forte, como também se especializa no subtópico de *causal inference* (inferência causal), refletindo um aprofundamento na busca por relações de causa e efeito, um dos maiores desafios da Economia.

Por fim, a evolução temática também aponta para a crescente autoconsciência do campo. A emergência do tópico *bibliometric analysis* no período recente é um forte indicativo de maturidade, pois demonstra que a área atingiu um volume e uma relevância tais que ela própria se tornou um objeto de estudo científico, exatamente como este trabalho se propõe a fazer. Essa transição geral, de ferramentas para problemas e de métodos para aplicações contextuais, ilustra um campo de pesquisa em plena desenvolvimento, adaptando-se continuamente para responder a questões econômicas cada vez mais complexas.

tes em grandes coleções de textos, baseado em modelos de distribuição probabilística.

5 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo central mapear o panorama científico da aplicação de *Machine Learning* (ML) no campo da Economia, utilizando para isso as ferramentas quantitativas da análise bibliométrica. A investigação partiu do reconhecimento da crescente proeminência da Ciência de Dados como uma força transformadora na teoria e na prática econômica. A análise realizada confirma esta percepção e revela um retrato detalhado de um campo de pesquisa em estado de desenvolvimento: marcado por um crescimento exponencial e recente, dotado de uma estrutura intelectual bem definida, mas também caracterizado por uma notável concentração geográfica e por uma colaboração científica que ainda se organiza em clusters isolados.

A síntese dos resultados aponta para uma história de maturação acelerada. Após um longo período latente, o interesse pela interseção entre ML e Economia explodiu a partir de 2017, impulsionando a produção científica a uma taxa de crescimento anual superior a 21%. A estrutura intelectual deste campo se organiza em torno de um núcleo metodológico, cujos artigos seminais constroem a ponte entre a predição algorítmica e a inferência econométrica. Este núcleo alimenta, principalmente, duas grandes frentes de aplicação: uma, já consolidada e motora do campo, em Finanças Quantitativas; e outra, um nicho especializado e em forte expansão, na Economia da Sustentabilidade. Social e geograficamente, a pesquisa é liderada por um ecossistema de instituições de elite predominantemente localizadas nos Estados Unidos, enquanto a comunidade de pesquisadores, apesar de altamente colaborativa, opera em redes fragmentadas, com pouca conexão entre os principais grupos de pesquisa.

Diante deste panorama, as contribuições deste estudo são multidimensionais. Para pesquisadores e estudantes que adentram a área, este trabalho oferece um mapa detalhado do território, identificando os autores, artigos e periódicos fundamentais, bem como as fronteiras de pesquisa mais ativas. Para instituições de fomento e departamentos acadêmicos, a análise evidencia a importância de cultivar ambientes interdisciplinares que unam a ciência da computação e a economia, ao mesmo tempo em que destaca as disparidades globais na produção de conhecimento, apontando para a necessidade de incentivar Redes de colaboração mais inclusivas e diversificadas.

É imperativo, contudo, reconhecer as limitações inerentes a esta análise. A metodologia bibliométrica, por sua natureza quantitativa, mede o impacto por meio de métricas de publicação e citação, não avaliando o mérito qualitativo intrínseco de cada contribuição. Ademais, a escolha das bases de dados Scopus e WoS, embora abrangente, implica um viés em favor de publicações em língua inglesa, podendo deixar de fora importantes contribuições regionais ou em outros idiomas. Outro ponto a ser considerado são as possíveis interpretações alternativas dos dados de colaboração e das redes temáticas, que podem refletir tanto estratégias de publicação quanto tendências momentâneas do campo. Essas limitações abrem avenidas para pesquisas futuras. Investigações qualitativas, como revisões sistemáticas aprofundadas sobre os *clusters* temáticos identificados (Finanças e Sustentabilidade), poderiam complementar os achados deste estudo. Além disso, análises futuras poderiam expandir o escopo para incluir outras bases de dados, anais de conferências e literatura em outras línguas, bem como explorar as causas subjacentes à fragmentação das redes de colaboração. Reconhecer essas restrições reforça a transparência científica e contribui

para um entendimento mais abrangente e crítico da dinâmica do campo.

Em última análise, este trabalho quantifica e caracteriza uma transformação em pleno andamento. A integração do ML não representa apenas a adição de uma nova ferramenta ao arsenal dos economistas, mas sim a emergência de um novo paradigma de pesquisa, capaz de extrair conhecimento de um mundo cada vez mais digital e rico em dados. Compreender a estrutura e a dinâmica deste campo não é apenas um exercício acadêmico, mas um passo essencial para orientar seu desenvolvimento futuro em direção a uma ciência econômica mais robusta, preditiva e relevante para os complexos desafios do século XXI.

REFERÊNCIAS

- ARIA, M.; CUCCURULLO, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, Elsevier, v. 11, n. 4, p. 959–975, 2017. 2, 11, 12
- ATHEY, S.; IMBENS, G. W. Machine learning methods that economists should know about. *Annual Review of Economics*, Annual Reviews, v. 11, p. 685–725, 2019. 4, 5
- ATHEY, S.; LUCA, M. Economists (and economics) in tech companies. *Journal of Economic Perspectives*, American Economic Association, v. 33, p. 209–230, dec 2019. ISSN 08953309. 2, 25
- BITETTO, A.; CERCHIELLO, P.; MERTZANIS, C. Measuring financial soundness around the world: A machine learning approach. *International Review of Financial Analysis*, Elsevier Inc., v. 85, jan 2023. ISSN 10575219. 6
- BROADUS, R. N. Toward a definition of “bibliometrics”. *Scientometrics*, Springer, v. 12, p. 373–379, 1987. 8
- CARMONA, P.; CLIMENT, F.; MOMPARDLER, A. Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, v. 61, p. 304–323, may 2019. 26
- CASABIANCA, E. J. et al. A machine learning approach to rank the determinants of banking crises over time and across countries. *Journal of International Money and Finance*, Elsevier Ltd, v. 129, dec 2022. ISSN 02615606. 6
- CHOLLET, F.; ALLAIRE, J. J. *Deep learning with R*. [S.l.]: Manning Publications, 2017. 2
- Clarivate. *Web Of Science*. [S.l.], 2024. Acesso em 26 de Fevereiro de 2024. Disponível em: <<https://www.webofscience.com/wos/>>. 9, 11, 14, 16, 17, 18, 19, 21, 22, 23, 25, 28, 29, 31, 32, 34
- COBO, M. J.; LÓPEZ-CÓZAR, E. D.; HERRERA-VIEDMA, E. *Bibliometrix: A Comprehensive R Package for Bibliometric Analysis*. [S.l.], 2019. Disponível em: <<https://cran.r-project.org/web/packages/bibliometrix/bibliometrix.pdf>>. 11, 12
- ECONOMIST, T. *The world's most valuable resource is no longer oil, but data*. 2017. <<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>>. Acesso em: 26 de maio de 2023. 1
- ELSEVIER. *Scopus: A Database of Academic Journals and Articles*. [S.l.], 2024. Acesso em 26 de Fevereiro de 2024. Disponível em: <<https://www.scopus.com/>>. 9, 11, 14, 16, 17, 18, 19, 21, 22, 23, 25, 28, 29, 31, 32, 34

- FINZER, W. The Data Science Education Dilemma. *Technology Innovations in Statistics Education*, California Digital Library (CDL), v. 7, 2013. 1
- GHODDUSI, H.; CREAMER, G. G.; RAFIZADEH, N. Machine learning in energy economics and finance: A review. *Energy Economics*, v. 81, p. 709–727, jun 2019. 26
- GU, S.; KELLY, B.; XIU, D. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, v. 33, n. 5, p. 2223–2273, may 2020. 26
- JESUS, D. P. de; BESARRIA, C. d. N. Machine learning and sentiment analysis: Projecting bank insolvency risk. *Research in Economics*, v. 77, n. 2, p. 226–238, 2023. ISSN 1090-9443. 5
- KESSLER, M. M. Bibliographic coupling between scientific papers. *American Documentation*, v. 14, n. 1, p. 10–25, 1963. 12
- KHAN, M. A. et al. Corporate vulnerability in the US and China during COVID-19: A machine learning approach. *Journal of Economic Asymmetries*, Elsevier B.V., v. 27, jun 2023. ISSN 17034949. 6
- KLEINBERG, J. et al. Human decisions and machine predictions. *The Quarterly Journal of Economics*, v. 133, n. 1, p. 237–293, feb 2018. 26
- KOZAK, S.; NAGEL, S.; SANTOSH, S. Shrinking the cross-section. *Journal of Financial Economics*, v. 135, n. 2, p. 271–292, feb 2020. 26
- LECHNER, M. Causal Machine Learning and its use for public policy. *Swiss Journal of Economics and Statistics*, Springer Science and Business Media Deutschland GmbH, v. 159, dec 2023. ISSN 22356282. 2, 4
- LEVINE, D. K. Economics: Eyes on the prize? *Science*, v. 323, p. 1296–1297, 2009. ISSN 00368075. 1
- MACIAS-CHAPULA, C. A. O papel da informetria e da ciênciometria e sua perspectiva nacional e internacional. *Ciência da Informação*, v. 27, n. 2, p. 134–140, 1998. 8
- MANELA, A.; MOREIRA, A. News implied volatility and disaster concerns. *Journal of Financial Economics*, v. 123, n. 1, p. 137–162, jan 2017. 26
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. In: *Sistemas Inteligentes Fundamentos e Aplicações*. 1. ed. Barueri-SP: Manole Ltda, 2003. p. 89–114. ISBN 85-204-168. 2
- MORETTIN, P. A.; SINGER, J. d. M. *Estatística e ciência de dados*. 1. ed. Rio de Janeiro: LTC, 2022. 2
- MULLAINATHAN, S.; SPIESS, J. Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, American Economic Association, v. 31, n. 2, p. 87–106, 2017. 25

NIKOU, M.; MANSOURFAR, G.; BAGHERZADEH, J. Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, v. 26, n. 4, p. 166–175, 2019. 26

PETROVÁ, K. The impact of digital technologies on Neoclassical Labour Market. *DANUBE*, v. 13, n. 4, p. 318–330, 2022. Disponível em: <<https://doi.org/10.2478/danb-2022-0020>>. 4

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>. 2, 11

RAUTENBERG, S.; CARMO, P. R. V. do. Big data e ciência de dados. *Brazilian Journal of Information Science: research trends*, Faculdade de Filosofia e Ciências, v. 13, p. 56–67, mar 2019. 1

SAGRADA, B. *Nova Versão Internacional (NVI)*. [S.l.], 2001. Disponível em: <<https://www.bibliaonline.com.br/nvi/sl/37/5-9>>. Acesso em: 24 jun. 2025. 4

VARIAN, H. R. Big data: New tricks for econometrics. *Journal of Economic Perspectives*, American Economic Association, v. 28, n. 2, p. 3–28, 2014. 25

ÍNDICE REMISSIVO

Acoplamento bibliográfico, 12

Cluster, 22, 23, 28–30, 32, 36

Coautoria, 2, 12, 13, 15, 20, 22, 31

Dilúvio de dados, 1

Machine Learning, 14–16, 18–20, 24, 27,
28, 30, 33, 35, 36, 38

Modelagem Computacional, 16

País, 6, 19, 23, 30–32

Políticas públicas, 4, 26

Produção Científica, 1, 14, 15, 20, 21, 30,
36

Redes de colaboração, 20, 24, 37

Índice H, 16, 20