



Avanço da Ciência de Dados nos Estudos de Economia: Uma análise bibliométrica do uso de *machine learning*

Caio Oliveira Azevedo

Fábio Junior Clemente Gama

Bruno Castro Alves

Kaylane Manuele Nunes Feitoza

RESUMO

O presente trabalho objetiva explorar os resultados de um mapeamento das pesquisas acadêmicas voltadas à Ciência de Dados nos estudos em Economia, com foco específico no uso de técnicas de *machine learning*. Busca identificar de que forma que a Ciência de Dados intercepta a economia e seus diversos eixos de atuação. Trata-se de uma pesquisa de cunho descritivo, em que foram realizadas análises bibliométricas por meio do pacote *Bibliometrix* do *software* R, de informações sobre produção acadêmica extraídas da base de dados da Scopus (SCP), tomando como recorte temporal o período de 2013 a 2023. Para a realização das análises foram utilizados os descritores “machine” e “learning”, tendo sido encontrados 1.608 trabalhos. Com base nos resultados foi possível confirmar a forte evolução na publicação de trabalhos que versam sobre *machine learning* aplicados a economia ou fatores econômicos, assim como o mapeamento dos principais eixos de discussões sobre o tema e as respectivas redes de autoria.

Palavras-chave: Ciência de dados, Aprendizado de máquina, Bibliometria, Economia.

1 INTRODUÇÃO

Ao longo das últimas décadas, têm sido observados avanços tecnológicos que alteraram a forma como a sociedade vive, tanto no âmbito profissional quanto na comunicação interpessoal. No cerne dessa revolução encontra-se a internet, um sistema global de informações que conecta pessoas, empresas e instituições em diferentes partes do mundo. A revolução tecnológica amparada, até então, pelas tecnologias de informação e comunicação, provocou um novo arranjo nas esferas da sociedade, sobretudo na forma de produzir ciência (LEVINE, 2009).

No final do século passado, os avanços tecnológicos elevaram a velocidade e o volume das transações de informação a um patamar surpreendente (FINZER, 2013), somados a maior capacidade de coleta e armazenamento de dados (*data warehousing*), proporcionaram um aumento drástico na quantidade de dados disponíveis e também de possibilidades de tomada de decisão orientada por dados. Configura-se, a partir de então, uma dinâmica de produção científica e de atuação profissional, pautada, principalmente, na maior magnitude do uso de recursos computacionais.

A massificação do uso das tecnologias de informação e comunicação, atrelada ao armazenamento de dados mais econômicos e ao processamento mais rápido realizado por computadores deu origem ao fenômeno



denominado por “dilúvio de dados”. Este fenômeno compreende um complexo sistema de dados extraídos das conexões entre as redes computacionais que envolvem o mundo contemporâneo (RAUTENBERG; CARMO, 2019).

O maior volume de informações provenientes da abundância de dados tem proporcionado aos pesquisadores científicos maior capacidade de aprofundar suas pesquisas. E não raro em alguns casos, possibilita avanços que antes eram inviáveis. Nesta perspectiva, é notório que o “dilúvio dos dados” tem transformado o processo de produção científica e sobretudo contribuído para avanços da exploração do conhecimento nas mais diversas áreas que interceptam o uso de dados (CLEVELAND, 2014).

Neste cenário, em que os dados e suas informações se tornaram um dos recursos mais valiosos (ECONOMIST, 2017), a Ciência de Dados (ou *Data Science* termo em inglês muito utilizado) emerge como uma abordagem multidisciplinar, configurando-se como um quarto paradigma da ciência, principalmente por ter se estabelecido como uma disciplina fundamental para a análise e interpretação de grandes conjuntos de dados (*big data*) em diversas áreas.

A Ciência de Dados compreende, deste modo, como um conjunto de suporte metodológico que combina princípios e práticas das áreas de matemática, estatística, inteligência artificial e engenharia da computação, direcionados para a captura, exploração e mineração de dados, a serem processados por softwares (CLEVELAND, 2014).

Tornar-se evidente, diante disso, que a Ciência de Dados é inerentemente também inter-disciplinar (FINZER, 2013). Uma simples consulta na internet é suficiente para constatar novos cursos de graduação ou centros de pós-graduação em Ciência de Dados em várias universidades ao redor do mundo, com programas de mestrado ou doutorado, implementados em faculdades de Ciência da Computação, Administração, Economia, Engenharia, Bioestatística etc.

Na área da Economia, por exemplo, essa abordagem tem desempenhado um papel cada vez mais frequente, oferecendo novas perspectivas e ferramentas para compreender a complexidade dos sistemas econômicos, principalmente através de *machine learning* (aprendizado de máquina ou aprendizado automático). Para Jordan (2019), o que é hoje rotulado como inteligência artificial, nada mais é do que aquilo que é chamado de *machine learning*.

Ferramentas como *machine learning* introduzem práticas de modelagem estatística em sistemas computacionais com a finalidade de desenvolver algoritmos capazes de instruir e adquirir conhecimentos de maneira automática (MONARD; BARANAUSKAS, 2003). Informalmente, a inteligência artificial está relacionada aos esforços de automatizar tarefas cognitivas humanas (CHOLLET; ALLAIRE, 2017), esta, portanto, visceralmente relacionada ao desenvolvimento de computadores ou de programação.

Nesta perspectiva convém ressaltar que o objetivo do *machine learning* não é o mesmo daquele considerado na análise de regressão, método tradicional em estudos econométricos. Enquanto na



econometria se pretende entender como cada variável preditora está associada com a variável resposta, no *machine learning* o objetivo é selecionar o modelo que produz melhores previsões, mesmo que as variáveis selecionadas com essa finalidade não sejam aquelas consideradas numa análise padrão (MORETTIN; SINGER, 2022).

Athey e Imbens (2019) destacam que a intersecção entre *machine learning* e os métodos tradicionais de econometria apresenta resultados de pesquisas muito mais produtivos, seguros e sofisticados. Os autores reforçam que a utilização de *machine learning* merece destaque na área de pesquisa econômica, uma vez que implica em avanços significativos.

Lechner (2023), por exemplo, relaciona os desdobramentos das transformações do uso da ciência de dados nos métodos econométricos e como esses mecanismos computacionais revolucionaram a robustez dos resultados, e por consequência das metodologias. Destaca que a utilização de algoritmos de *machine learning* é capaz de obter resultados muito mais eficientes sobre os impactos de uma política pública e seus efeitos nos subgrupos e público-alvo.

É evidente, portanto, que a Ciência de Dados intercepta a economia e seus diversos eixos de atuação. Nesta perspectiva, este artigo tem como objetivo explorar o avanço da Ciência de Dados nos estudos em Economia, com foco específico no uso de técnicas de *machine learning*. Para tanto, propõe-se uma abordagem bibliométrica, que permitirá examinar e quantificar o progresso e a evolução desses estudos ao longo do tempo, bem como mensurar as colaborações entre autores de diversos países, uma vez que essas métricas são indicativas do grau de expansividade e repercussão da temática analisada.

A análise bibliométrica é amplamente reconhecida como uma ferramenta valiosa para avaliar a produção científica, identificar tendências emergentes e analisar a colaboração entre pesquisadores em um determinado campo. Neste estudo, será aplicada uma série de métricas bibliométricas, como o número de publicações, a frequência de citações e a identificação de redes de coautoria, para examinar a produção científica relacionada ao uso de *machine learning* em Economia.

O objetivo deste estudo bibliométrico é contribuir para a compreensão do avanço da Ciência de Dados nos estudos de Economia, fornecendo um panorama das tendências e abordagens predominantes. Além disso, espera-se que as métricas bibliométricas utilizadas ofereçam *insights* valiosos para pesquisadores, profissionais e estudantes interessados em explorar as aplicações do *machine learning* nesse campo.

Para esta finalidade, utilizou-se um método quantitativo de medição estatística, com aplicações no *software R*¹ e o uso da interface *Biblioshiny* do pacote *bibliometrix*². As buscas foram realizadas na base de

¹ R Core Team (2022)

² Aria e Cuccurullo (2017)



dados da Scopus (SCP), tomando como recorte temporal o período de 2013 a 2023. Para a realização das análises foram utilizados os descritores “*machine*” e “*learning*”, tendo sido encontrados 1.415 trabalhos.

Com base nos resultados foi possível confirmar a forte evolução na publicação de trabalhos que versam sobre *machine learning* aplicados a economia ou fatores econômicos, também o mapeamento dos principais eixos de discussões sobre o tema e as respectivas redes de autoria.

Além desta introdução, este artigo se divide em mais 4 seções. A próxima 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 *machine learning* 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

Esta seção apresenta estudos seminais referentes à temática do presente artigo, além de contextualizar o recente histórico de desenvolvimento e contribuição de estudos empíricos acerca da eficiência na utilização de modelos de *machine learning* aplicados às Ciências Econômicas.

Petrová (2022) afirma que o advento da revolução tecnológica impactou as esferas do mercado de trabalho neoclássico, tendo em vista que essas ferramentas, como ressalta o autor, são fatores imprescindíveis para fortalecer a economia dos países. Para acompanhar as transformações causadas pelo progresso tecnológico, a dinâmica dos novos economistas no mercado deve ser ampliada pela qualificação no uso desses mecanismos. É diante desse cenário que trabalhos da área de economia têm evidenciado a integração entre técnicas de aprendizado de máquinas e métodos tradicionais de análises econômicas.

A incorporação dos sistemas inteligentes tem marcado profundamente as novas dinâmicas de trabalho da economia como evidencia-se diante das pesquisas com aplicações na econometria, que apresentam avanços para a compreensão do comportamento de relações causais de variáveis que são foco de estudo dessa ciência (LECHNER, 2023). Nesse sentido, o arranjo de ferramentas do economista deve ser expandido em conformidade ao uso de sistemas de *Machine Learning*, de modo a preservar, simultaneamente, os principais métodos tradicionais da econometria (ATHEY; IMBENS, 2019).

O ponto crucial da ciência econômica é a habilidade de interpretar as informações estabelecidas em um conjunto de dados, desenvolvida a partir do uso de modelagem estatística a fim de apresentar um panorama de atuação em âmbitos sociais, políticos e privados, além de otimizar e orientar as tomadas de decisões. Diante disso, o estudo intitulado *Causal Machine Learning and its use for public policy*, publicado pela revista norte-americana *Swiss Journal of Economics and Statistics* de Lechner (2023) aponta o exponencial desenvolvimento tecnológico dos sistemas inteligentes e a relevância de aplicações de técnicas de *Causal Machine Learning* no campo econômico.



Os resultados provenientes permitem estimativas mais robustas e precisas, além do ponto diferencial do uso dessas técnicas em relação aos métodos econométricos tradicionais que é identificar as heterogeneidades dos impactos dessas políticas, de modo a investigá-las metodicamente com a finalidade de alcançar os objetivos previstos durante a elaboração dessas ações, as aperfeiçoando e corrigindo as lacunas identificadas, e assim usar os recursos públicoseficientemente (LECHNER, 2023).

Conforme apresentam Athey e Imbens (2019), uma das principais e inovadoras aplicações dos métodos de ML é para a estimação de efeitos médios de tratamento, um parâmetro essencial para as análises econômicas, sejam elas de níveis macro ou microeconômicos. Tais métodos apresentam desempenhos mais eficientes em relação aos métodos econométricos tradicionais, principalmente quando se trata de analisar grandes bases de dados ou mecanismos mais complexos da distribuição de tratamento. Os autores do estudo apresentam outro cenário de usadas técnicas do aprendizado de máquina este com foco nas estimações de políticas ótimas em modelos econômicos. Os métodos de ML para estimativas ótimas de políticas são capazes de

propiciar os pesquisadores a levar em consideração um conjunto de opções ao captar a complexidade das relações entre as variáveis, com a finalidade de promover sugestões de políticas bem mais precisas (ATHEY; IMBENS, 2019).

Além da utilização de engrenagens de sistemas inteligentes em estudos econométricos, no campo da macroeconomia esses mecanismos também tem sido incorporados nas análises de estimações e causalidades de agregados macroeconômicos. de Jesus e Besarria (2023) desenvolveram a partir do uso de técnicas de *K-means*, um método não supervisionado de ML que processa *clusters* da amostra das instituições bancárias, a montagem de novas métricas para a classificação de riscos de insolvência bancária para as instituições financeiras que negociam em bolsas brasileiras. O emprego dessas técnicas apresenta diferenciais nos resultados por possuírem a capacidade de extrair os dados não estruturados e tratar a não-linearidade deles de maneira eficaz (de Jesus; BESARRIA, 2023).

Em conformidade a esse cenário, Casabianca et al. (2022) ao utilizarem o modelo *logit* tradicional e um algoritmo de aprendizado de máquina supervisionado, (*AdaBoots*), para identificar quais os determinantes de crises bancárias entre as amostras de países em estudo e concluíram que o desempenho do mecanismo de *Machine Learning* sobressai o modelo tradicional. Esse aspecto foi observado pelos autores a partir da análise feita fora da amostra em que o modelo preditivo de *AdaBoots* apresentou desempenho mais eficiente diante da amostra, uma vez que essa ferramenta se destaca em termos de "AUROC, sensibilidade, especificidade e utilidade relativa" (CASABIANCA et al., 2022).

Bitetto, Cerchiello e Mertzanis (2023) também elaboraram um estudo acerca da estimativa de causalidade no crescimento econômico e políticas expansionistas com uma abordagem técnica de modelagem de elevação (*uplift*), uma área ainda em exploração do aprendizado de máquina. A estrutura de ML mencionada permite



trabalhar com efeito causais, estimando a causalidade em cada indivíduo sobre o crescimento real do PIB e mudanças de política econômica expansionistas. Esta análise se concentra no uso de algoritmos causais que são ramificados com a finalidade de identificar efeitos causais defasados nas políticas expansionistas e, diante dessa identificação, estimar o impacto no crescimento econômico (BITETTO; CERCHIELLO; MERTZANIS, 2023). Os autores salientam que a modelagem *Uplift* apresenta resultados que além de irem ao encontro da literatura, também expressam comportamentos macroeconômico da teoria e da realidade.

No cenário de corporações e finanças, o emprego de *Machine Learning*, como apontam os autores Khan et al. (2023), promove uma previsão mais concisa acerca do nível de vulnerabilidade corporativa da amostra de países estudados ao identificar relações entre os efeitos da pandemia COVID-19 e esse cenário de crise nas instituições financeiras. Apesar dessa defesa, o estudo argumenta que a modelagem de ML pode apresentar risco a precisão dos resultados pelas suposições processadas pelo algoritmo, dado a análise da performance dessa ferramenta em previsão de risco de default de crédito apresentar riscos ao método de validação de supervisão (KHAN et al., 2023).

Os efeitos de mudanças do cenário sociopolítico no campo macro e microeconômico, estimativas e previsões acerca das implicações de políticas econômicas, são exemplos de análises econômicas que podem ser feitas e aprimoradas a partir do uso de sistemas inteligentes como *Machine Learning* em conjunto aos métodos tradicionais de análise de dados. Evidencia-se, portanto, nessa conjuntura de avanços da Ciência de Dados, que os algoritmos de ML auxiliam e potencializam as tomadas de decisões a partir dos resultados por eles obtidos, sobretudo aqueles direcionados para previsões, como apontam os estudos levantados.

3 PROCEDIMENTOS METODOLÓGICOS

3.1 ESTRATÉGIA EMPÍRICA

A bibliometria é um método quantitativo (medição estatística) para mapeamento e avaliação de uma área de pesquisa a partir dos dados bibliográficos de sua produção científica, possibilitando ao pesquisador identificar, desde os artigos mais citados da área aos termos e conceitos mais relevantes (SILVA; HAYASHI; HAYASHI, 2011; MEDEIROS; VITORIANO, 2015).

Neste sentido, no campo da bibliometria, a estratégia empírica é fundamental para coletar e analisar dados sobre a produção científica, inclui citações, autores, revistas e palavras-chave. Essa abordagem visa identificar tendências, padrões e áreas de pesquisa relevantes que contribuem para o desenvolvimento de um sólido corpo de conhecimento (BROADUS, 1987).

A estratégia empírica na bibliometria passa por várias etapas. Inicialmente, os pesquisadores coletam dados bibliográficos de fontes confiáveis, como bases de dados acadêmicas, periódicos científicos e conferências relevantes. Em seguida, esses dados são organizados e estruturados para facilitar a análise quantitativa e qualitativa.



Durante a análise empírica, são aplicadas técnicas estatísticas e de visualização de dados para identificar padrões na produção científica, como fontes mais citadas, principais autores, revistas de maior impacto e temas emergentes. Essas informações proporcionam uma compreensão abrangente do campo de estudo, permite a identificação de áreas de pesquisa promissoras, lacunas de conhecimento e oportunidades para futuras investigações.

Com essa finalidade, nas próximas duas subseções será aplicado o processo de coleta de dados, assim como do tratamento destes, de modo a explorar o panorama científico de forma sistemática e embasada em dados, contribui para o avanço do conhecimento e a promoção de discussões relevantes no campo de estudo em questão.

3.2 DADOS

O mapeamento bibliométrico a partir do uso de métricas estatísticas e matemáticas permite a elaboração e análise da revisão sistemática de literatura com a finalidade de identificar e avaliar os avanços da utilização de engrenagens computacionais de *Machine Learning* na economia nos últimos dez anos. Diante disso, os dados foram coletados da base da Scopus Elsevier (ELSEVIER, 2023).

A escolha da base de dados da Scopus parte das seguintes razões: (i) abrangência e diversidade de conteúdo; (ii) qualidade e precisão dos dados: a plataforma é conhecida por sua rigorosa seleção de fontes e processos de indexação, o que resulta em alta qualidade e precisão dos dados.; (iii) métricas de citação e impacto, fundamental para realização da análise de citação.

Além das observações ressaltadas, a base da Scopus Elsevier comporta parte significativa das principais produções científicas da área de economia e ciência de dados, o que possibilita um mapeamento bibliométrico de maneira mais ampla e concisa entre a interdisciplinaridade dos eixos abordados.

Foram registrados 1608 artigos ao utilizar o comando de pesquisa avançada da Scopus com os parâmetros a seguir: TITLE-ABS-KEY (machine AND learning) AND PUBYEAR

> 2012 AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (OA, "all")) AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (SUBJAREA, "ECON")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English")).

O critério de delimitação dos parâmetros no intervalo de tempo de dez anos parte da finalidade de garantir que os artigos pesquisados estejam atualizados em conformidade às evoluções da Ciência de Dados, sobretudo para identificar a tendência de potenciais avanços da utilização de *Machine Learning* em trabalhos econômicos diante da expansão do fenômeno do *Big Data* na última década. As restrições seguidas pelo comando “and” objetivam a seleção dos artigos provenientes de revistas acadêmicas que são relacionados à área específica da Economia e que estão publicados.



Tabela 1 – Descrição das variáveis pré-selecionadas

Variáveis Analisadas	Nomenclatura	Descrição
Annual Scientific Production	ASP	Refere-se ao acompanhamento da produção de documentos científicos em um campo de estudo específico ao longo do período de análise.
Most Global Cited Documents	MGCD	Refere-se aos documentos científicos mais citados globalmente dentro de um determinado campo de estudo ou área de pesquisa. Esses documentos são aqueles que receberam o maior número de citações de outros trabalhos científicos em âmbito internacional, demonstrando sua ampla influência e reconhecimento na comunidade acadêmica.
WordCloud	WC	Conhecida como nuvem de palavras, é uma representação visual das palavras mais frequentes em um texto ou conjunto de textos. Nessa representação, as palavras são dispostas em tamanhos proporcionais à sua frequência, ou seja, palavras que aparecem mais vezes têm um tamanho maior na nuvem.
Co-occurrence Network	CON	Representa as relações entre elementos que coocorrem em um conjunto de dados ou documentos. Nesse tipo de rede, os elementos são representados por nós (ou vértices) e as coocorrências entre eles são representadas por arestas (ou conexões).
Collaboration Network	COLN	Representa as colaborações entre pesquisadores, países ou instituições em um determinado campo de pesquisa ou área de atuação.

Fonte: Elaboração própria a partir de Elsevier (2023).

Diante disso, como mostra a Tabela 1, as variáveis analisadas neste estudo bibliométrico desempenham um papel fundamental na compreensão e avaliação da pesquisa científica. Essas variáveis combinadas oferecem uma abordagem ampla e valiosa para explorar a pesquisa científica, permitindo assim uma análise aprofundada e uma visão mais completa das tendências, interações e impacto da produção acadêmica nesse campo em constante evolução.

3.3 TRATAMENTO

O tratamento dos dados coletados para a análise bibliométrica foi realizado a partir do *software* R³, em que utilizou-se *bibliometrix*⁴, pacote que habilita a interface entre aplicativo *biblioshiny*. Este é uma ferramenta que possibilita a geração e visualização de diversas análises bibliométricas em arquivos de dados bibliográficos (ARIA; CUCCURULLO, 2017).

De acordo com Hjørland (2008), o acoplamento bibliográfico foi introduzido por Kessler (1963), e a análise de co-citação foi sugerida independentemente por Marshakova e Small, ambos em 1973. Em seus experimentos, Kessler (1963) encontrou um alto grau de correlação semântica entre os documentos agrupados ao utilizarem o critério de acoplamento bibliográfico.

³ R Core Team (2022)

⁴ Cobo, López-Cózar e Herrera-Viedma (2019)



Os pacotes de clusterização utilizam o acoplamento bibliográfico como base para seu processo. O bibliometrix emprega o algoritmo de clusterização "armadilha da caminhada", método eficaz para agrupar documentos bibliográficos com base em suas citações. O algoritmo seleciona um documento aleatório e, em seguida, expande-se para agrupar outros que são citados pelo inicial. Esse processo é repetido até que todos os documentos tenham sido agrupados. O resultado é um conjunto de clusters de documentos que são semanticamente relacionados (ARIA; CUCCURULLO, 2017).

Uma rede de acoplamento pode ser obtida a partir da formulação geral:

$$B = A \times A^T \quad (1)$$

onde A é uma rede bipartida. O elemento b_{ij} indica quantos acoplamentos bibliográficos existem entre os manuscritos i e j . Em outras palavras, b_{ij} fornece o número de caminhos de comprimento 2, que se move de i ao longo da seta e, em seguida, para j na direção oposta. A matriz B é simétrica ($B = B^T$). A força do acoplamento de dois artigos, i e j , é definida simplesmente pelo número de referências que os artigos têm em comum, conforme fornecido pelo elemento b_{ij} da matriz B .

O grau de acoplamento bibliográfico entre A e B (Rede bipartida) é determinado pela frequência com que seus documentos são citados simultaneamente, como é o caso de C .

E para construir a rede de colaboração por clusters de autores, utiliza-se a seguinte formulação geral:

$$AC = A^T \times A \quad (2)$$

onde: A é uma rede bipartida Manuscritos ~~X~~ Autores. Como a matriz B , a matriz C também é simétrica. A diagonal principal de C contém o número de casos em que uma referência é citada em nosso quadro de dados. Em outras palavras, o elemento diagonal c_i é o número de citações locais da referência i .

Ao final desse processo, em que as matrizes de dados já formadas pelo bibliometrix são indexadas na interface, inicia-se a execução do mapeamento científico, através da utilização dos principais indicadores bibliométricos para identificar as métricas do tema estudado.

Ao analisar o corpus bibliográfico selecionado, buscaremos identificar padrões de publicação, principais autores e instituições, colaborações, periódicos mais relevantes e palavras-chave mais recorrentes. Essas análises nos permitirão avaliar o crescimento da literatura científica nessa área específica, do mesmo modo identificar os principais tópicos e abordagens de pesquisa adotados pelos estudiosos da interseção entre Ciência de Dados e Economia.

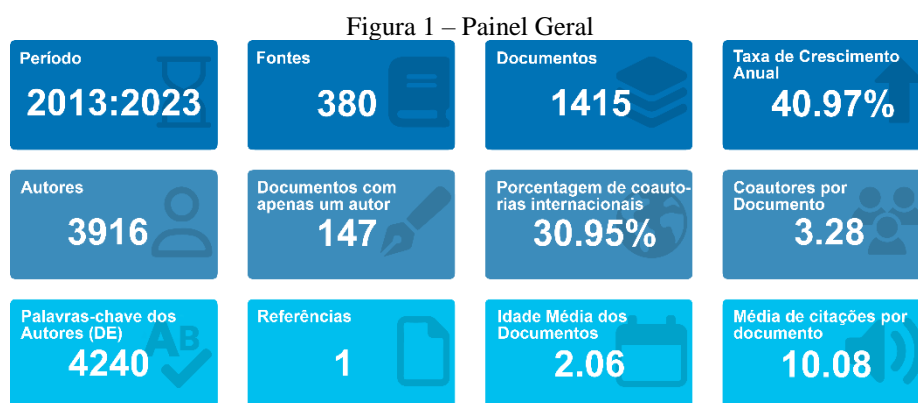


4 RESULTADOS

A revisão sistemática da literatura utilizou como base 1.415 documentos de pesquisa obtidos a partir de 380 fontes diferentes, publicados entre 2013 e 2023. Os dados revelam uma taxa de crescimento anual significativa de 40,97% que demonstra os avanços da aplicação dessas ferramentas no âmbito econômico, longo dos últimos dez anos. Além disso, os documentos apresentam uma idade média de 2,06 anos, parâmetro que evidencia a atualidade dessa temática e seu crescimento em conformidade aos avanços da ciência de dados.

O número médio de citações por documento de 10,08 indica que o conteúdo destes documentos tem recebido notoriedade e reconhecimento científico na comunidade acadêmica. Outro fator analisado é o conjunto de dados o qual envolve um total de 3.916 autores que contribuíram para os 1.415 documentos, demonstrando uma tendência de colaboração entre autores tendo uma média de 3,28 coautores por documento.

Em contrapartida, 147 dos documentos são de autoria de um único indivíduo o que representa 10% do total analisado. Isso sugere que a pesquisa em conjunto, no cenário do *Big Data*, é uma prática relevante nessa nova dinâmica de produção científica marcada pela intersecção da Ciência de Dados e a Economia ao permitir perspectivas e influências intelectuais mais amplas diante da aplicação de sistemas inteligentes para a elaboração das pesquisas. Além disso, aproximadamente 30,95% das colaborações envolvem parcerias internacionais, indicativa da abrangência do tema nas pesquisas acadêmicas para além das fronteiras das instituições dos países identificados.



Fonte: Elaboração própria a partir de Elsevier (2023).

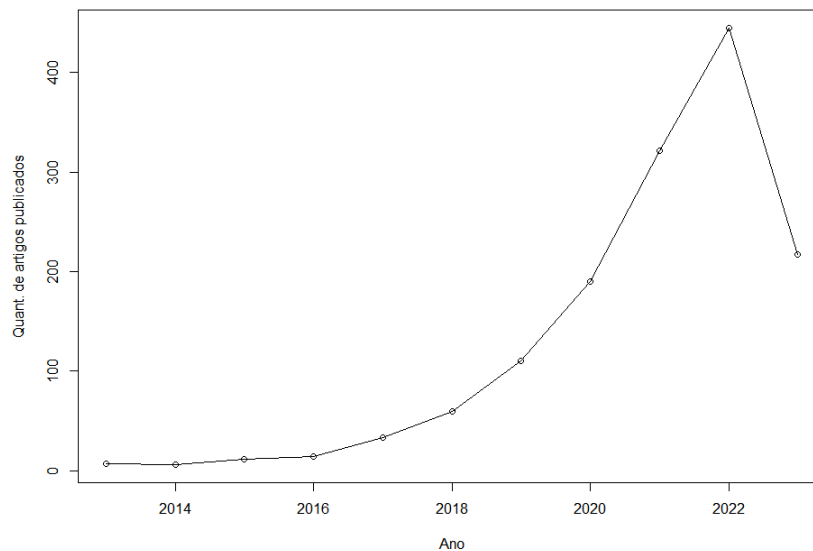
A trajetória de crescimento nas publicações de artigos sobre *machine learning* na economia ao longo dos anos tem um comportamento exponencial (Figura 2), ratificando a tendência da intersecção entre Ciência de Dados e Teoria Econômica. Inicialmente, observou-se um crescimento constante e moderado até 2016, seguido por um aumento acelerado. Isso está relacionado principalmente aos avanços tecnológicos, somados a maior capacidade de coleta e armazenamento de dados (*data warehousing*), diante da era do dilúvio de dados, proporcionando maiores possibilidades na aplicação de métodos de inteligência artificial.



Esse cenário de contínuo avanços tecnológicos proporciona aos pesquisadores maior facilidade de retenção, obtenção e armazenamentos de dados variados que, ao serem tratados mediante a aplicação de algoritmos de ML, podem ser utilizados como matéria-prima para desenvolvimento de pesquisas cada vez mais aprofundadas. Essas inovações possibilitam a resolução de problemáticas em trabalhos que anteriormente eram pouco exploradas em decorrência da ausência de informações contidas nesses registros cibernéticos e da dificuldade de extraí-las e armazená-las. Nesse sentido, os resultados cada vez mais eficientes de trabalhos voltados para essa temática impulsionam as produções de artigos nessa área, o que explica o avanço expressivo de 2018 até os anos seguintes, confirmando-se, assim, as alterações otimistas da nova dinâmica de produção científica diante dos avanços das tecnologias em espaços da ciência econômica.

É importante destacar que a análise de 2023 levou em conta exclusivamente os seis primeiros meses do ano, parâmetro que influenciou nos totais apresentados. Assim, é esperado que a tendência de expansão de trabalhos publicados, que continuou em um trajeto crescente até 2022, alcançando o pico com 445 publicações neste ano, seja mantida para os meses seguintes, ao considerar a magnitude dos números já registrados nos seis primeiros meses, que não só conseguem superar os doze meses de 2021 como também aproximam-se dos resultados do ano completo de 2022.

Figura 2 – Distribuição dos artigos ao longo dos anos
Artigos sobre Machine Learning na Economia, publicados a partir de 2013



Fonte: Elaboração própria a partir de Elsevier (2023).

Nesta perspectiva, a Tabela 2 apresenta os trabalhos mais citados pelos artigos que versam sobre *machine learning*. A análise de citação fornece as principais referências teóricas sobre o campo em questão e as contribuições dos periódicos para a evolução do tema. As associações conceituais com maior incidência evidenciam o quanto essas produções são relevantes para o campo de estudo da economia com inserção de



métodos de ML, uma vez que indicam o suporteargumentativo para o desenvolvimento de trabalhos nessa área.

O primeiro entre os mais citados, *Double/debiased machine learning for treatment and structural parameters*, aborda o desafio da inferência em um parâmetro de baixa dimensionalidade (θ_0) diante de parâmetros de alta dimensionalidade (η_0). Os autores introduzem o conceito de "Double or Debiased Machine Learning" (DML) e propõem um método que utiliza técnicas de aprendizado de máquina e momentos ortogonais de Neyman para estimar θ_0 de forma consistente, de modo a superar o viés introduzido pela regularização e *overfitting* na estimação de η_0 (CHERNOZHUKOV et al., 2018).

Na sequência, o segundo trabalho mais citado é *Empirical Asset Pricing via Machine Learning*, que consiste numa análise comparativa de métodos de aprendizado de máquina para o problema clássico de precificação empírica de ativos. Os autores demonstram que o uso de previsões de aprendizado de máquina, em específico, de árvores de decisão e redes neurais, podem gerar grandes ganhos econômicos para os investidores. O aumento preditivo obtido, frente aos modelos de regressão, é atribuído a capacidade dos métodos de *machine learning* em capturar interações não lineares entre os preditores, o que não é explorado pela regressões lineares (GU; KELLY; XIU, 2020).

O artigo *Machine learning and deep learning* é o terceiro mais citado, com 342 citações e sendo o mais recente entre os três primeiros, publicado em 2021. O trabalho explora a utilização de sistemas inteligentes com capacidades de inteligência artificial, frequentemente baseados em aprendizado de máquina e aprendizado profundo. O estudo proporciona uma visão conceitual desses métodos, com ênfase tanto nos desafios enfrentados ao implementar tais sistemas em mercados eletrônicos e negócios em rede, quanto na importância da interação humano-máquina e da prestação de serviços de inteligência artificial (JANIESCH; ZSCHECH; HEINRICH, 2021). A tabela 3 permite visualizar a diversificação das áreas dos periódicos que lideram a publicação de artigos no campo em questão, com uma clara tendência para a área de finanças. Esse resultado reforça o uso de sistemas inteligentes para melhores performances de trabalhos em análises de riscos e vulnerabilidade no setor financeiro. Destaque para os periódicos *Quantitative Finance* e *Financial Innovation*, ambos de grande relevância internacional, com forte influência nos centros de pesquisas e de ensino acadêmico.

Tabela 2 – Tabela Documentos mais citados pelos artigos que versam sobre Machine learning

os.	Documentos	Citações	Periódico
	Chernozhukov et al. (2018)	520	<i>The Econometrics Journal</i>
	Gu, Kelly e Xiu (2020)	369	<i>The Review of Financial Studies</i>
	Janiesch, Zschech e Heinrich (2021)	342	<i>Electronic Markets</i>
	Dubey et al. (2020)	216	<i>International Journal of Production Economics</i>
	Baek, Mohanty e Glambosky (2020)	211	<i>Finance Research Letters</i>



Huang e Rust (2021)	206	<i>Journal of the Academy of Marketing Science</i>
Athey e Imbens (2019)	202	<i>Annual Review of Economics</i>
Deryugina et al. (2019)	171	<i>American Economic Review</i>
Zhu et al. (2019)	156	<i>International Journal of Production Economics</i>
Belloni et al. (2017)	149	<i>Econometrica</i>

0

Fonte: Elaboração própria a partir de Elsevier (2023).

A Figura 3 apresenta os termos mais frequentes em artigos científicos da aplicação de métodos de *machine learning* no eixo econômico. O objetivo principal desta análise é identificar as tendências e áreas de maior interesse nesse campo de pesquisa. A nuvem de palavras, extraída das *keywords*, aponta os principais métodos de ML aplicados nesses estudos. A partir dessa análise, constatou-se o que o termo *forecasting* aparece com maior predominância, demonstrando a tendência da aplicação de técnicas de algoritmos de ML para fins de previsão.

Figura 3 – Distribuição de palavras mais frequentes nos artigos



Fonte: Elaboração própria a partir de Elsevier (2023).

A ocorrência, também em evidência, de *learning algorithms* sugere que os estudos estão concentrados na investigação e comparação de diferentes algoritmos de aprendizado de máquina. Esse enfoque destaca a importância do desenvolvimento de abordagens algorítmicas eficientes para solucionar problemas específicos dos diversos eixos da economia, uma vez que o desempenho eficiente de métodos de *machine learning* está intimamente relacionado não só com as estruturas dos dados trabalhados, mas com os objetivos esperados pela aplicação, que confirma os algoritmos mais adequados de acordo com cada particularidade de resultados que são buscados no desenvolvimento das pesquisas.

Seguido deste, o termo *learning systems* reflete a importância atribuída à implementação prática de algoritmos de aprendizado de máquina em sistemas reais e abre a possibilidade de variações na nomenclatura mencionada nos artigos. Em relação às técnicas específicas de aprendizado de máquina, os elementos *decision trees* e *support vector machines* são recorrentes, o que indica o interesse em explorar esses modelos para classificação e maximização de desempenho preditivos que auxiliam em tomadas de decisões assertivas fundamentadas em resultados mais robustos.



Tabela 3 – Periódico que mais publicaram artigos que versam sobre Machine learning

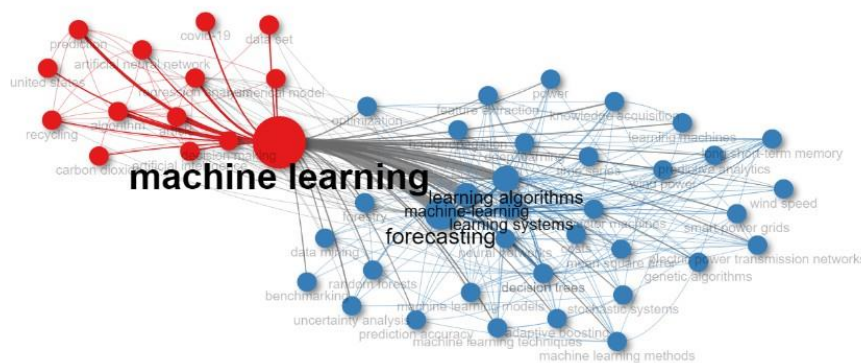
Fontes	Artigos	Qualis	Área
<i>Frontiers in Energy Research</i>	107	A3	Biotecnologia
<i>Journal of Advanced Transportation</i>	77	A2	Administração Pública e de Empresas, Ciências Contábeis e Turismo
<i>Risks</i>	76	B2	Economia
<i>Forecasting</i>	34	A2	Economia
<i>Humanities and Social Sciences Communications</i>	29	B4	Ciências Sociais
<i>Journal of Risk and Financial Management</i>	29	B2	Economia
<i>Journal of Finance and Data Science</i>	28	-	Finanças e Ciência de Dados
<i>Quantitativa Finance</i>	27	A1	Economia
<i>Financial Innovation</i>	19	A1	Administração Pública e de Empresas, Ciências Contábeis e Turismo
<i>Periodicals of Engineering and Natural Sciences</i>	19	-	Engenharia

Fonte: Elaboração própria a partir de Elsevier (2023).

Ademais, os termos *algorithm* e *deep learning* também são encontrados com certa frequência, indicando uma análise abrangente e aprofundada sobre o desenvolvimento e a aplicação de diversos algoritmos, como são os casos de redes neurais profundas para aprender padrões complexos. Por fim, o termo *regression analysis* também é destacado, o que revela o interesse em analisar a relação entre variáveis por meio de modelagem estatística, aplicada em contextos particulares de ML.

Além disso, vale observar que dentre as palavras mais utilizadas nos artigos, basicamente estas estão divididas em dois subgrupos, como apontado pela análise de Coocorrência de palavras. A análise da Rede de Coocorrência de palavras-chave, como mostrado na Figura 4, aponta a existência de clusters temáticos distintos. Clusters são conjuntos de objetos ou dados que compartilham características semelhantes, constituem em uma técnica importante na análise e organização de informações. O processo de formação dos clusters envolve o uso de técnicas de agrupamento (clustering), as quais analisam as similaridades entre os elementos e os agrupam com base nessas características, criando grupos coesos e distintos. O cluster vermelho é composto por pesquisas de Machine Learning voltadas para análise de regressão e sua associação com a tomada de decisão. Esse agrupamento permite explorar como técnicas de regressão têm sido aplicadas no contexto da tomada de decisões, seja em sistemas autônomos ou em ambientes de negócios complexos. Por outro lado, o Cluster Azul abrange artigos que abordam diversas técnicas de aprendizado de máquina, com enfoque em previsão e Modelos de Machine Learning. Nesse cluster, pode-se encontrar trabalhos que exploram algoritmos preditivos, como árvores de decisão, redes neurais, máquinas de suporte vetor, entre outros, bem como estudos que tratam da criação e avaliação de diferentes modelos de aprendizado de máquina para resolver problemas específicos em diversas áreas.

Figura 4 – Rede de Coocorrência por meio de palavras-chave



Fonte: Elaboração própria a partir de Elsevier (2023).

Como destacado, cerca de 90% das publicações relacionadas ao tema são compostas por mais de um autor, portanto, podemos concluir que há redes de cooperação entre os autores. O bibliometrix permite avaliar essas redes através de gráficos de cooperação entre países e autores. A figura 5 apresenta análise das parcerias internacionais entre países no campo de pesquisa econômica por meio do Mapa Mundial da Colaboração dos Autores em artigos sobre *machine learning*. Os resultados indicam que os Estados Unidos (USA) e o Reino Unido lideram as colaborações, com 37 parcerias, seguidos pela cooperação entre os Estados Unidos e a China (28 colaborações), e China e Reino Unido (23 colaborações).

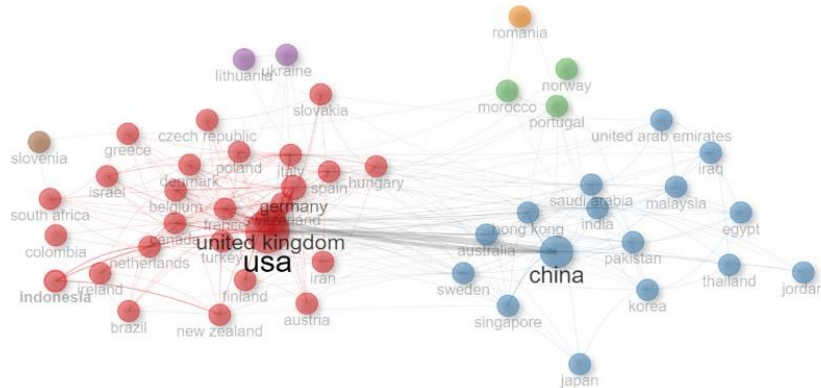
A análise de colaboração é uma ferramenta importante para estudos bibliométricos. Ela permite identificar redes de pesquisadores que trabalham em conjunto, bem como as áreas de pesquisa que estão mais ativas e colaborativas. Isso é útil para identificar tendências de pesquisa, rastrear o desenvolvimento de novas áreas e identificar pesquisadores que estão fazendo contribuições importantes para um campo.

Através dessas parcerias, é possível enfrentar desafios complexos, desenvolver novas abordagens e aprimorar as técnicas de aprendizado de máquina. Esses resultados destacam a importância da cooperação internacional para o progresso científico e tecnológico, permite avanços significativos no campo do *machine learning*. As parcerias entre países contribuem para o desenvolvimento de soluções inovadoras e podem impactar positivamente a sociedade, abre caminho para avanços em áreas como medicina, indústria, finanças, entre outras.

A colaboração de diferentes autores em trabalhos científicos é essencial para impulsionar a pesquisa acadêmica uma vez que a natureza complexa e interdisciplinar dos problemas contemporâneos exige a combinação de expertise e conhecimentos de múltiplos especialistas. Através da colaboração os autores podem agregar e combinar suas perspectivas, habilidades e experiências acadêmicas individuais, de modo a enriquecer a qualidade e a abrangência dos trabalhos científicos. Ao analisar o gráfico 6, destaca-se o Cluster

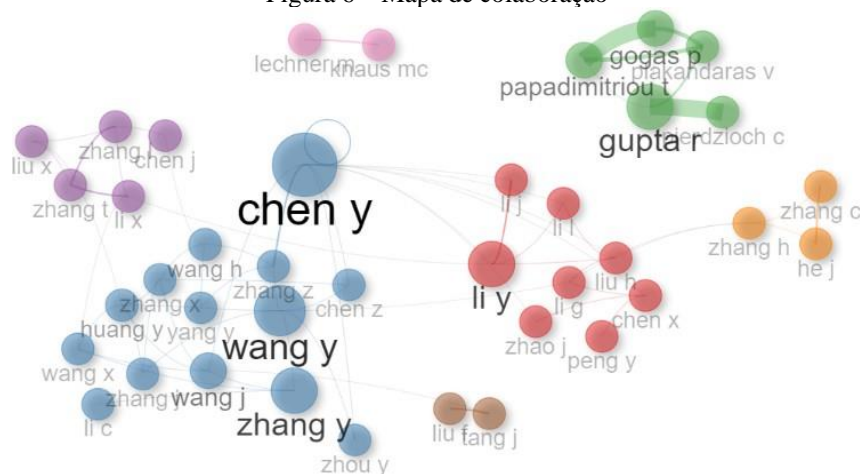
Azul que é composto por autores como Chen Y⁵, Wang Y⁶, Zhang Y⁷, Wang H⁸, Wang J⁹ e dentre outros. Dentre eles, o autor Chen Y se destaca com um valor bastante elevado em "Entrelinhas", sugere uma significativa contribuição para os artigos em machine learning. Nesses clusters, observamos uma rede densa de colaboração, evidenciada pelos valores relativamente altos de "Proximidade" para todos os autores, indica que eles colaboram estreitamente entre si na produção dos artigos. "

Figura 5 – Colaboração entre autores de diferentes países



Fonte: Elaboração própria a partir de Elsevier (2023).

Figura 6 – Mapa de colaboração



Fonte: Elaboração própria a partir de Elsevier (2023).

O crescimento exponencial evidenciado nessa análise reflete o reconhecimento cada vez maior da relevância do *machine learning* no aprofundamento da compreensão e aprimoramento dos processos econômicos. Por meio de análises avançadas e decisões embasadas em dados, o ML desponta como uma poderosa ferramenta capaz de impulsionar a eficiência e a eficácia dos negócios, catalisa inovações e

⁵ Ver Chen (2023)

⁶ Ver Liang et al. (2023)

⁷ Ver Sasaki, Ura e Zhang (2022)

⁸ Ver Gao, Wang e Gao (2021)

⁹ Ver Wang, Tang e Guo (2022)



descobertas que moldarão de maneira significativa o panorama econômico do futuro. Seu contínuo desenvolvimento é de suma importância para enfrentar os desafios e aproveitar as oportunidades de um cenário global em constante transformação tecnológica.

5 CONSIDERAÇÕES FINAIS

É reconhecido o avanço da Ciência de Dados no campo da Teoria Econômica ao longo da última década devido às evidências da abordagem multidisciplinar dessa ciência em conformidade ao aumento da relevância dos dados como recursos decisivos para as tomadas de decisões (ECONOMIST, 2017), sobretudo a partir da utilização de algoritmos de aprendizado de máquina.

A análise bibliométrica desenvolvida no presente trabalho comprova a elevada incidência do interesse e engajamento do campo econômico no uso de ferramentas de *machine learning*, ao apresentar um crescimento constante até 2016, como aponta a figura 2, seguido por um rápido aumento no número de publicações. Esse resultado sugere que o pesquisador das Ciências Econômicas, no cenário de destaque do *Big Data*, é conduzido a aplicar algoritmos que o auxiliam na análise de grandes bases de dados. Através do uso de análises avançadas e tomadas de decisão informadas, o *machine learning* tem o potencial de impulsionar a eficiência e eficácia dos negócios e promover inovações no futuro econômico.

Outro resultado fornecido por esta bibliometria pode ser analisado a partir dos três artigos mais citados, destacados na tabela 2, os quais demonstram a diversidade de tópicos relevantes em *machine learning* que estão sendo aplicados em estudos econômicos. Os levantamentos desses estudos apontam o desenvolvimento de novos arranjos técnicos de aprendizado de máquina e suas aplicações práticas em finanças, confirmando, a partir da análise das revistas que mais utilizam algoritmos de ML, o destaque dessa área no uso dessas ferramentas para otimizar o desempenho preditivo de análises de riscos encontrados diante da volatilidade de mercados financeiros. A extração e mineração de informações dos dados permite a esses pesquisadores ampliarem suas análises acerca das movimentações financeiras ao redor do mundo em decorrência da magnitude dos dados cada vez mais abrangente.

As análises também fornecem informações sobre as áreas de foco e tendências predominantes, pontuando o uso de técnicas como "Forecasting", "Decision Trees" e "Support Vector Machines". Essas conclusões indicam a aplicação de métodos de ML para análises preditivas, modelagens e mapeamentos de resultados e decisões. Na prática, essa aplicação demonstra o interesse e a responsabilidade dos pesquisadores por resultados mais sofisticados e robustos no que se refere a previsões de impactos ou variações econômicas nos diversos eixos da economia, entre outras conjunturas desse campo que exigem tomadas de decisões eficientes.

Outro aspecto que o estudo revela é o avanço expressivo da trajetória dessa temática com relação ao intercâmbio teórico, como é observado pela ampla rede de colaboração entre pesquisadores de diferentes



localidades do mundo. Essa análise permite compreender que a liderança de parcerias internacionais parte do proeminente arcabouço acadêmico de países que apresentam essencialmente determinada consolidação em desenvolvimentos tecnológicos, como Estados Unidos, Reino Unido e China, cenário de nova dinâmica de produção científica potencializada pela busca por inovações capazes de aprimorar a abordagem de dados na economia. Desse modo, evidencia-se que a aplicação de engrenagens computacionais de ML avança em conformidade à trajetória de ascendência das tecnologias.

Em síntese, as contribuições provenientes dessa pesquisa fornecem direcionamentos teórico e prático para a exploração das principais tendências da utilização de ML em notoriedade na Teoria Econômica. A partir da bibliometria, pesquisadores e profissionais interessados no eixo abordado pelo estudo podem visualizar as principais finalidades das aplicações de algoritmos dos sistemas inteligentes para abordagens semelhantes às identificadas durante a pesquisa.

Além disso, o estudo apresenta contribuições teóricas para pesquisas futuras na área de economia e inovação, conforme os resultados que confirmam a interseção da Ciência de Dados e a Economia. Os parâmetros avaliados durante o estudo comprovam as tendências ascendentes da aplicação de ML, fornecendo, assim, *insights* pertinentes para a comunidade acadêmica desenvolver e impulsionar trabalhos com o suporte eficiente de algoritmos de *Machine Learning*, uma vez que o desenvolvimento de campos científicos é uma trajetória gradual e contínua.

CLASSIFICAÇÃO JEL

A12, B00, B49, C55



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. Disponível em: <<https://doi.org/10.1016/j.joi.2017.08.007>>.
- 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.
- BAEK, S.; MOHANTY, S. K.; GLAMBOSKY, M. Covid-19 and stock market volatility: An industry level analysis. *Finance Research Letters*, v. 37, 2020.
- BELLONI, A. et al. Program evaluation and causal inference with high-dimensional data. *Econometrica*, v. 85, n. 1, p. 233 – 298, 2017.
- 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, 1 2023. ISSN 10575219.
- BROADUS, R. N. Toward a definition of “bibliometrics”. *Scientometrics*, Springer, v. 12, p. 373–379, 1987.
- CASABIANCA, E. J. et al. A machine learning approach to rank the determinants of banking crises overtime and across countries. *Journal of International Money and Finance*, Elsevier Ltd, v. 129, 12 2022. ISSN 02615606.
- CHEN, Y. Comparing content marketing strategies of digital brands using machine learning. *Humanities and Social Sciences Communications*, v. 10, n. 1, 2023.
- CHERNOZHUKOV, V. et al. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, v. 21, n. 1, p. C1–C68, 01 2018.
- CHOLLET, F.; ALLAIRE, J. J. *Deep learning with R*. [S.l.]: Manning Publications, 2017.
- CLEVELAND, W. S. Data science: An action plan for expanding the technical areas of the field of statistics. *Statistical Analysis and Data Mining*, John Wiley and Sons Inc, v. 7, p. 414–417, 12 2014. ISSN 19321872.
- 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>>.
- de Jesus, D. P.; BESARRIA, C. da N. Machine learning and sentiment analysis: Projecting bank insolvency risk. *Research in Economics*, v. 77, n. 2, p. 226–238, 2023. ISSN 1090-9443.
- DERYUGINA, T. et al. The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, v. 109, n. 12, p. 4178 – 4219, 2019.
- DUBEY, R. et al. Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. *International Journal of Production Economics*, v. 226, 2020.



ECONOMIST, T. *The world's most valuable resource is no longer oil, but data* (2017). 2017. Disponível em: <<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>>.

ELSEVIER. *Scopus: A Database of Academic Journals and Articles*. [S.l.], 2023. Acesso em 10 de maio de 2023. Disponível em: <<https://www.scopus.com/>>.

FINZER, W. The data science education dilemma. *Technology Innovations in Statistics Education*, California Digital Library (CDL), v. 7, 2013.

GAO, G.; WANG, H.; GAO, P. Establishing a credit risk evaluation system for smes using the soft voting fusion model. *Risks*, v. 9, n. 11, 2021.

GU, S.; KELLY, B.; XIU, D. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, v. 33, n. 5, p. 2223–2273, 02 2020.

HJORLAND, B. What is knowledge organization (ko)? *Knowledge Organization*, v. 35, n. 2/3, p. 86–101, 2008.

HUANG, M.-H.; RUST, R. T. A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, v. 49, n. 1, p. 30 – 50, 2021.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. *Electronic Markets*, Springer, v. 31, n. 3, p. 685–695, 2021.

JORDAN, M. I. Artificial intelligence—the revolution hasn't happened yet. *Harvard Data Science Review*, PubPub, v. 1, n. 1, p. 1–9, 2019.

KESSLER, M. M. Bibliographic coupling between scientific papers. *American Documentation*, v. 14, n. 1, p. 10–25, 1963.

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, 6 2023. ISSN 17034949.

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, 12 2023. ISSN 22356282.

LEVINE, D. K. Economics: Eyes on the prize? *Science*, v. 323, p. 1296–1297, 2009. ISSN 00368075.

LIANG, D. et al. A novel fault monitoring method based on impedance estimation of power line communication equipment. *Frontiers in Energy Research*, v. 11, 2023.

MEDEIROS, J. M. G. de; VITORIANO, M. A. V. A evolução da bibliometria e sua interdisciplinaridade na produção científica brasileira. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, v. 13, n. 3, p. 491–503, 2015.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, v. 1, n. 1, p. 32, 2003.

MORETTIN, P. A.; SINGER, J. d. M. *Estatística e ciência de dados*. 1ª. ed. Rio de Janeiro: LTC, 2022.

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>>.



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

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, 3 2019.

SASAKI, Y.; URA, T.; ZHANG, Y. Unconditional quantile regression with high-dimensional data. *Quantitative Economics*, v. 13, n. 3, p. 955 – 978, 2022.

SILVA, M. R. da; HAYASHI, C. R. M.; HAYASHI, M. C. P. I. Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. *InCID: revista de ciência da informação e documentação*, v. 2, n. 1, p. 110–129, 2011.

WANG, J.; TANG, J.; GUO, K. Green bond index prediction based on ceemdan-lstm. *Frontiers in Energy Research*, v. 9, 2022.

ZHU, Y. et al. Forecasting smes' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. *International Journal of Production Economics*, v. 211, p. 22 – 33, 2019.