

# ESTUDO BIBLIOMÉTRICO: ML APLICADA NA AGRICULTURA

Data de submissão:

Data de aceite: 01/08/2023

### Gizele Ferreira da Silva

Mestranda PGDRA - Fundação  
Universidade Federal de Rondônia.

### David Lopes Maciel

Doutorando PGDRA – UFRO. Professional  
Master in Master of Science in Emergent  
Technologies in Education. MUST  
UNIVERSITY, MUST, EUA.

### Fabrcio Moraes de Almeida

PhD in Physics (UFC), with post-doctorate  
in Scientific Regional Development (DCR/  
CNPq). Researcher of the Doctoral and  
Master Program in Regional Development  
and Environment (PGDRA/UFRO).

**RESUMO:** A análise estatística de informações bibliográficas são a base para os estudos bibliométricos e a concepção de modelos ou leis que tratam do desenvolvimento do conhecimento no estado da arte. No século XIX, surge de forma incipiente a primeira expressão mais sistemática, entretanto só no início do século XX, com a publicação dos trabalhos de *Lotka* ela ganha força com inserção dos indicadores de produção (BOUSTANY, 1997). A Bibliometria faz uso de métodos matemáticos afim de descrever e quantificar

estudos relacionados a uma temática científica (PRITCHARD, 1969). Neste contexto, o objetivo deste paper é quantificar as publicações na área de *Machine Learning (ML)* aplicada na agricultura, através de uma análise bibliométrica. Para tanto, utilizou-se do banco de dados obtido na base *Scopus* e *Web of Science*. O processo de análises fez uso do *software R / RStudio* e da aplicação *Bibliometrix* e sua biblioteca *Biblioshiny*, a partir dos dados, verificou-se que o maior número de publicações sobre o tema ocorreu nos anos 2021 e 2022, o autor que mais publicou foi o *WANG Y*, os periódicos de maior relevância foram os *COMPUTERS AND ELECTRONICS IN AGRICULTURE* e *REMOTE SENSING*.

**PALAVRAS-CHAVE:** *Machine learning*.  
Aplicação. Agricultura.

## BIBLIOMETRIC STUDY: ML APPLIED IN AGRICULTURE

**ABSTRACT:** The statistical analysis of bibliographic information is the basis for bibliometric studies and the conception of models or laws that deal with the development of knowledge in the state of the art. In the 19th century, the first more systematic expression appears in

an incipient way, however only in the beginning of the 20th century, with the publication of Lotka's works, does it gain strength with the insertion of production indicators (BOUSTANY, 1997). Bibliometrics makes use of mathematical methods in order to describe and quantify studies related to a scientific theme (PRITCHARD, 1969). In this context, the objective of this paper is to quantify the publications in the area of Machine Learning (ML) applied in agriculture, through a bibliometric analysis. For that, we used the database obtained from Scopus and Web of Science. The analysis process made use of the R / RStudio software and the Bibliometrix application and its Biblioshiny library, from the data, it was verified that the largest number of publications on the subject occurred in the years 2021 and 2022, the author who most published was the WANG Y, the most relevant journals were COMPUTERS AND ELECTRONICS IN AGRICULTURE and REMOTE SENSING.

**KEYWORDS:** Machine Learning. Application. Agriculture.

## 1 | INTRODUÇÃO

Este capítulo de livro trata da análise da produção científica mundial sobre a aplicação da inteligência artificial na agricultura por meio de indicadores bibliométricos. O tema foi escolhido por estar em evidência e apresentar grande relevância para o meio produtivo do agronegócio, configurando-se em uma área estratégica na produção de alimentos em nível mundial, neste contexto a melhoria e ampliação da produção de alimentos é de interesse geral, vem de encontro ao objetivo dois da agenda 2030.

Para a realização do estudo, foi usado a bibliometria, que é um método de análise quantitativa para a pesquisa científica. Os dados elaborados por meio dos estudos bibliométricos estimam a contribuição do conhecimento científico para o meio acadêmico e a sociedade, derivando das bases de dados as publicações, e do conjunto delas os indicadores utilizados nas análises e discursões (SOARES, CARNEIRO, et al., 2016).

Ainda de acordo com a bibliometria, que é definida como o estudo da mensuração e quantificação do progresso científico, estando a pesquisa baseada em indicadores bibliométricos (BIANCHI, 2002), o mapeamento científico usa métodos bibliométricos para examinar disciplinas, campos, especialidades, autores e redes de pesquisa, busca identificar como elas se relacionam entre si.

Desta forma, o método produz mapas e representações espacial e temporal dos dados obtidos nas bases. Corrobora com esta definição, *Boyack e Klavans* quando afirmam que “[...] O mapeamento científico é uma combinação de classificação e visualização” (BOYACK, 2005, p. 358). O objetivo é criar uma representação da estrutura da área de pesquisa, particionando elementos (documentos, autores, periódicos, palavras) em diferentes grupos. A visualização é então usada para criar uma representação visual da classificação que emerge. (ČATER, 2015).

Os *softwares* utilizados neste trabalho foram o R / RStudio / Bibliometrix / Biblioshiny em conjunto com o MS Excel, as bases utilizadas foram *Web of Science* e a *Scopus*, por

apresentarem o maior número de indexação de periódicos. A pesquisa foi sistematizada através do método SSF - *Systematic-Search Flow*, que tem como objetivo sistematizar o processo de busca, a fim de garantir a repetibilidade e a replicabilidade, evitando os vieses tendenciosos por parte do pesquisador (FERENHOF E FERNANDES, 2016).

## 2 | METODOLOGIA

Neste capítulo, a operacionalização da pesquisa se deu através da aplicação do método SSF, que consiste de 4 fases e 8 atividades, conforme apresentado na Fig. 1, o método foi desenvolvido segundo os autores, com o intuito de sistematizar o processo de busca ou buscas à base de dados científicas. Sendo assim, ele serve tanto para a revisão sistemática como para a revisão integrativa, dependendo apenas da definição da estratégia na sua utilização (FERENHOF E FERNANDES, 2016).

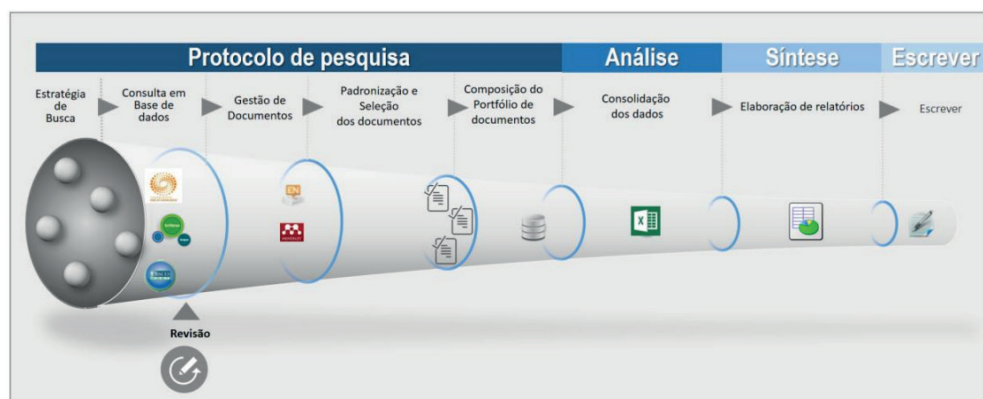


Figura 1 - Representação do método *Systematic Search Flow*

Fonte: Ferenhof e Fernandes, 2016.

Vale ressaltar, que por ocasião da utilização do software R<sup>®</sup>, na fase 1 (Protocolo de Pesquisa) as atividades 3, 4 e 5, foram totalmente automatizadas com o uso das Bibliotecas RStudio / Bibliometrix. Na fase 2 (Análise) a atividade (consolidação dos dados) também fora automatizada com as mesmas bibliotecas do R<sup>®</sup>, entretanto foi gerado um arquivo do tipo Banco de dados no MS Excel<sup>®</sup>, contendo todos os artigos pesquisados e unificando os dados obtidos na base *Web of Science* e *Scopus*, este serviu para tratar novamente os dados oriundo do *Bibliometrix*. A fase 3 (Síntese) seguiu o mesmo procedimento de automatização do processo com o uso da biblioteca / *Biblioshine* do R<sup>®</sup> aplicada no Banco de Dados gerado na fase anterior.

Já na fase 4, conforme preconizado pelos autores Ferenhof e Fernandes, realizou-se a consolidação dos resultados por intermédio da escrita científica, por tanto, foi resgatado o objetivo do trabalho, comparado com o resultado da análise e da síntese, nesta fase foi

eliminada a matriz do conhecimento recomendada pelos autores, uma vez que os dados foram gerados dentro do *Biblioshine*, bem como, os relatórios que fundamentaram a escrita dos resultados, finalizando a atividade 8, proposta pelos autores.

## 2.1 Execução da Pesquisa

Protocolo da pesquisa:

- Atividade 1 - estratégia de busca fazendo uso do operador lógico (AND, \*, “), e do Query de pesquisa (artigos, documentos em Inglês, artigos indexados no período entre 2015 e 2022, e aplicação de filtros nas categorias da Web of Science);
- Atividade 2 - parametrização e aplicação das Query de pesquisa na base selecionada conforme figuras 2, os termos utilizados foram (*machine learning, application e agriculture*).

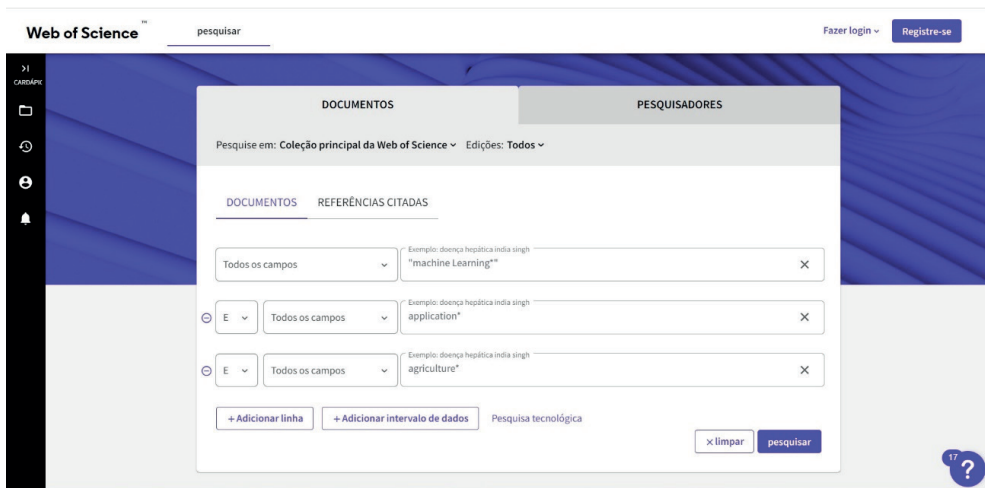
The image shows a screenshot of the Web of Science search interface. At the top, there is a search bar with the text "pesquisar" and a "Fazer login" button. Below the search bar, there are two tabs: "DOCUMENTOS" and "PESQUISADORES". The "DOCUMENTOS" tab is selected. The search query is displayed as "Pesquise em: Coleção principal da Web of Science" with "Edições: Todos" selected. The search results are displayed in a table with two columns: "DOCUMENTOS" and "REFERÊNCIAS CITADAS". The search query is "machine Learning\*" and "application\*" and "agriculture\*". There are three search criteria listed, each with a dropdown menu set to "Todos os campos" and a search button. The search criteria are: "machine Learning\*", "application\*", and "agriculture\*". There are also buttons for "+ Adicionar linha", "+ Adicionar intervalo de dados", and "Pesquisa tecnológica". A "limpar" button is also present.

Figura 2 - Parametrização da busca na Base de Dados da *Web of Science*.

Fonte: autores, 2023.

A pesquisa foi realizada na Base de dado da *Web of Science*, através do portal de periódico da CAPES – CAFe.

## 3 | RESULTADOS

A partir da busca inicial, foi gerado o banco de dado contendo as informações de artigos indexados na base *Web of Science* com 476 artigos, esta fora baixada na extensão “BiTex” conforme mostrado na figura 3 – Print do resultado da Pesquisa. Após download o banco foi exportado para o *Bibliometrix R*, onde passou por tratamento gerando um arquivo

MS Excel® da base. Novamente este arquivo passou por uma reanálise e tratamento onde o critério foi manter apenas os artigos que possuíam DOI, sendo eliminados também as duplicações, desta forma, restaram 475 artigos, sendo eliminados 1 artigo por não possuir DOI. Na sequência estes foram exportados para interface do R – *Biblioshiny*, onde foram gerados as informações e os diversos gráficos para análise posterior dos resultados.



Figura 3 – Print do resultado da Pesquisa

Fonte: autores, 2023.

A figura 3, apresenta os dados obtidos na base da *Web of Science*, contendo os 476 artigos, a figura apresenta ainda a configuração do query de pesquisa e os filtros aplicados.

### 3.1 Resultados do processamento no Biblioshiny.

Após o processamento dos dados na biblioteca R / *Bibliometrix*, foram obtidos os seguintes resultados apresentados nas figuras 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14.

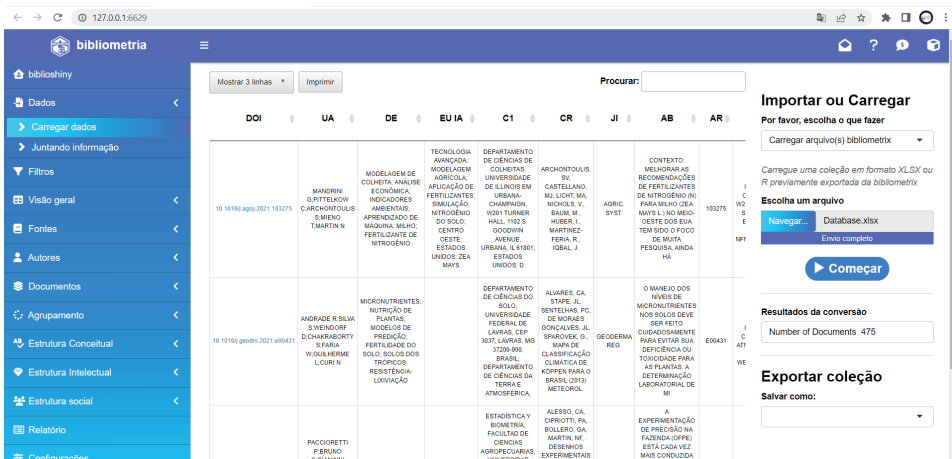


Figura 4 – interface de importação do *Biblioshiny*.

Fonte: autores, 2023.

Na interface de importação dos dados, é possível verificar os dados que foram obtidos e correlaciona-los com o arquivo base, tais como intervalo de tempo dos documentos selecionados, número de autores, referências e documentos, tipos de documentos, dentre outros.



Figura 5 – Main Information.

Fonte: autores, 2023.

A figura 5, apresenta em números as informações de forma geral, e um resumo das informações obtidas.

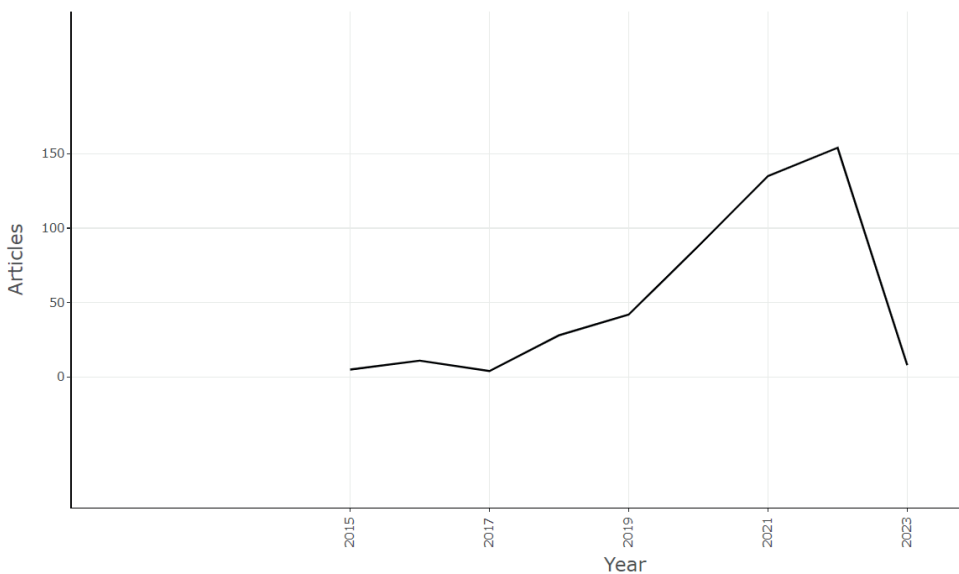


Figura 6 – Annual Scientific Production.

Fonte: autores, 2023.

A figura 6, apresenta as informações da produção científica anual.

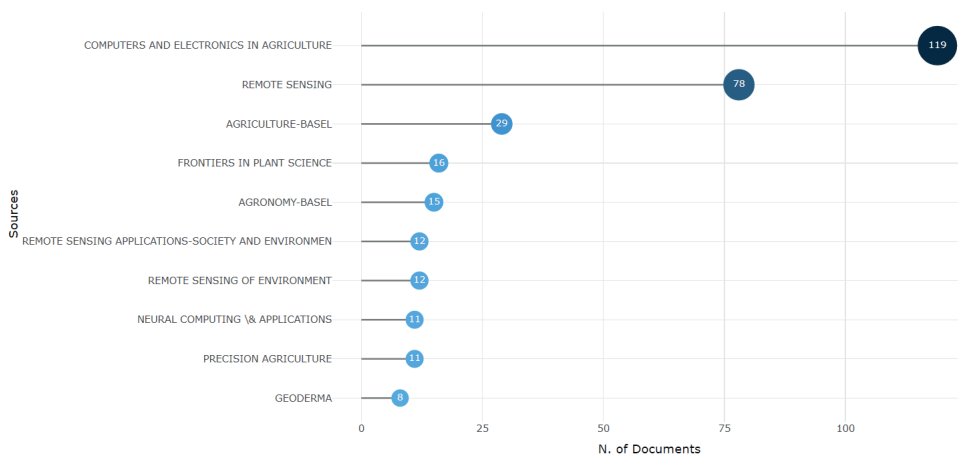


Figura 7 – *Most Relevant Sources*.

Fonte: autores, 2023.

Já a figura 7, traz o gráfico com a indicação das fontes mais relevantes em nível mundial.

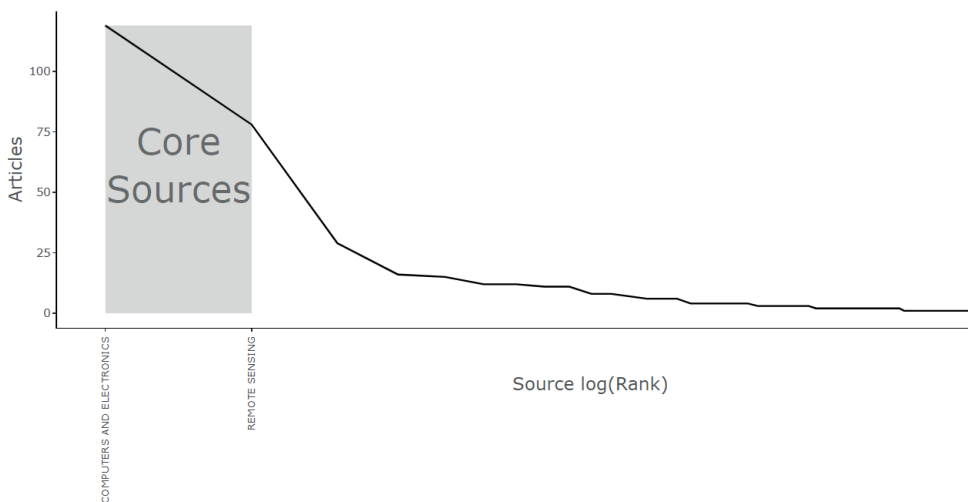


Figura 8 – *Core Sources by Bradford's Law*.

Fonte: autores, 2023.

Na figura 8, o gráfico apresenta as publicações levando em conta as principais fontes categorizadas pela Lei de *Bradford*. Esta lei dispõe sobre a ordenação decrescente da produtividade de artigos de determinado assunto nos periódicos científico, possibilitando

o estabelecimento de agrupamentos divididos de forma exponencial.

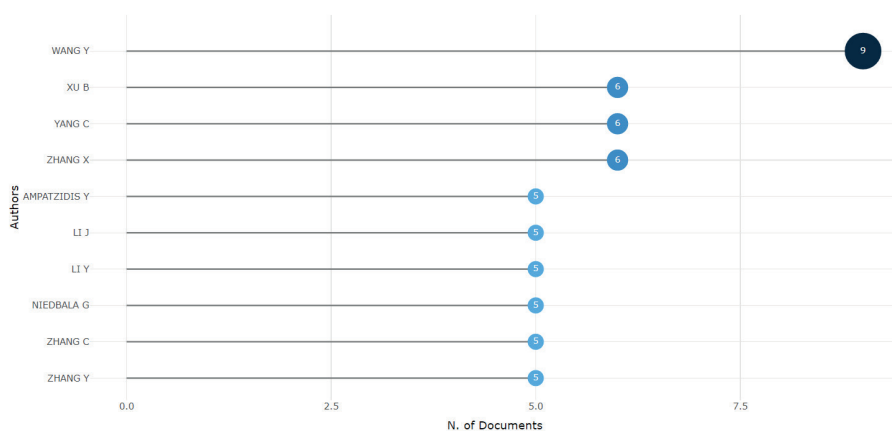


Figura 9 – *Most Relevant Authors.*

Fonte: autores, 2023.

A figura 9, apresenta o gráfico com os principais autores com maior relevância no cenário internacional.

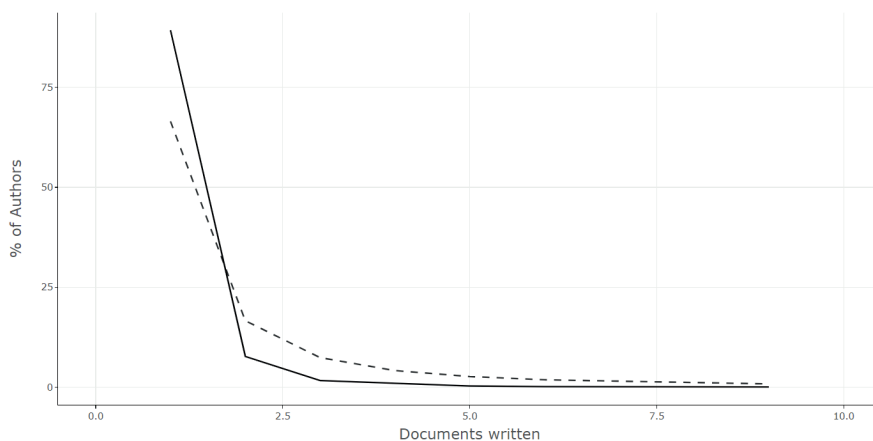


Figura 10 – *Author Productivity through Lotka's Law.*

Fonte: autores, 2023.

A figura 10, nos traz a produtividade dos autores analisada através da Lei de *Lotka*, esta lei descreve a frequência de publicação de autores em qualquer campo do conhecimento, com base no quadrado inverso, em que o número de autores que publicam um determinado número de artigos é uma proporção fixa para o número de autores que



publicam um único artigo.

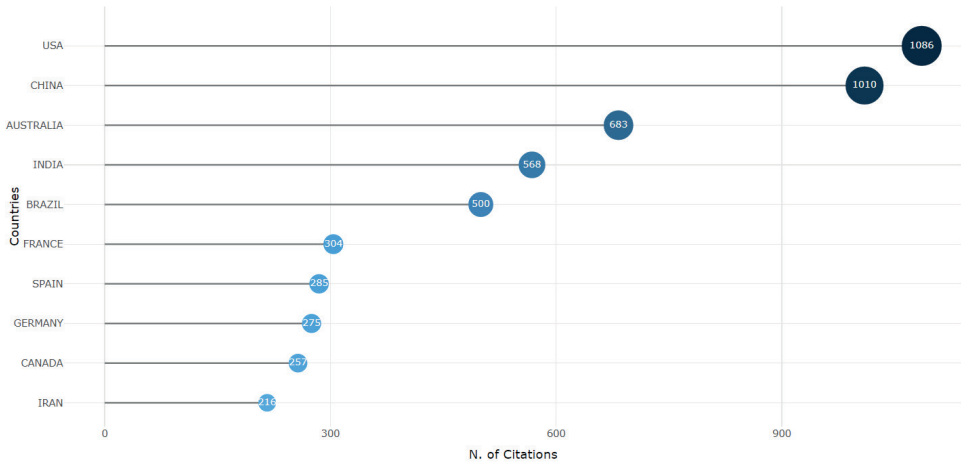


Figura 11 – Most Cited Countries.

Fonte: autores, 2023.

A figura 11, apresenta o gráfico os países mais citados nas publicações científicas e a frequência de citações.



Figura 12 – WordCloud.

Fonte: autores, 2023.

A figura 12, apresenta em forma de nuvem de palavras, os termos mais utilizados nas publicações, bem como a frequência de sua ocorrência. Quanto maior o tamanho da palavra, maior a frequência de utilização, logo, quanto menor o tamanho da palavra, menor sua frequência.

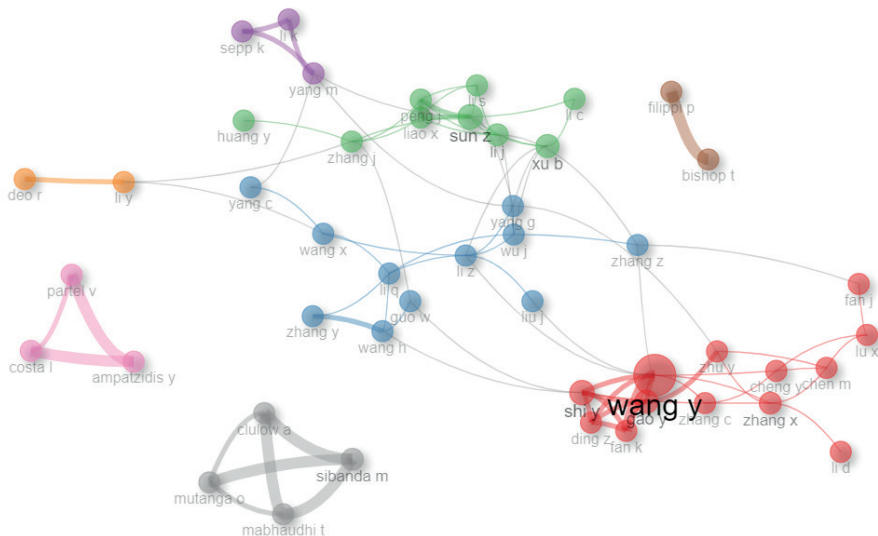


Figura 13 – *Collaboration Network*.

Fonte: autores, 2023.

Por fim, a figura 13, apresenta graficamente as redes internas de autores (sem ligação com outras redes), permitindo evidenciar quais são os autores e as redes que mais publicaram sobre o tema abordado.

## CONCLUSÃO

O uso dos indicadores bibliométricos para estudar as atividades de pesquisa de um ou mais campo do conhecimento, tem se tornado cada vez mais comum na avaliação da produção científica. Esses indicadores podem incluir informações sobre a quantidade de artigos publicados, a frequência de citações recebidas, o impacto das revistas em que foram publicados, entre outros dados.

Dessa forma, a bibliometria permite, portanto, uma análise quantitativa da produção científica, que pode ser útil tanto para a comparação entre diferentes áreas de conhecimento quanto para a avaliação individual de pesquisadores e instituições. No entanto, é importante lembrar que a bibliometria não deve ser utilizada como única fonte de avaliação, devendo ser complementada por outras metodologias qualitativas e subjetivas.

Além disso, do ponto de vista cognitivo, a inovação do conhecimento somente

adquire o seu valor quando ele é difundido dentro da comunidade científica. Isso ocorre porque a produção de conhecimento é um processo social, e é somente através da interação entre os pesquisadores que novas ideias e descobertas podem ser avaliadas, testadas e integradas ao conhecimento já existente (MICHAEL, 2000).

Isto leva a que se possa afirmar que a fidedignidade ou a confiabilidade dos resultados dos estudos bibliométricos irá depender substancialmente de sua aplicação correta, levando em conta suas vantagens, mas também as limitações e condições necessárias de sua utilização (ČATER, 2015).

Sendo assim, o estudo bibliométrico realizado, utilizando o método SSF de (FERENHOF E FERNANDES, 2016) e adaptado pelos autores do artigo, possibilitou a obtenção, tratamento e análise dos dados, bem como a geração dos resultados que hora passamos a analisar. Partindo das informações obtidas e apresentadas na figura 6, que o tema é atual e apresenta um padrão de crescimento constante, o que pode ser observado na figura 5, onde a taxa de crescimento é 6,05% ao ano, com idade média de 2,39 anos nos conjuntos dos documentos.

A figura 7, apresenta o gráfico com a indicação das fontes com maior revência global, é possível observar que em se tratando do tema, periódicos multidisciplinares englobam a maior parte dos artigos, sendo o caso do periódico *Computers and Electronics in Agriculture - Copyright © 2023 Elsevier B.V. All*, que fornece cobertura internacional de avanços no desenvolvimento e aplicação de hardware de computador, *software*, instrumentação eletrônica e sistemas de controle para resolver problemas na agricultura, incluindo agronomia e horticultura, se destaca com 119 publicações envolvendo o tema proposta.

A figura 8, apresenta o gráfico das publicações levando em conta a lei de *Bradford*, A *Lei de Bradford* que sugere que à medida que os primeiros artigos sobre um novo assunto são escritos, eles são submetidos a uma pequena seleção, por periódicos apropriados, e se aceitos, esses periódicos atraem mais e mais artigos, no decorrer do desenvolvimento da área/assunto. Ao mesmo tempo, outros periódicos publicam seus primeiros artigos sobre o assunto (BEUREN, 2007), o que pode ser confirmado no gráfico com o *Computers and Electronics in Agriculture*, seguido do periódico REMOTE SENSING- MODIS (*Moderate Resolution Imaging Spectroradiometer*), ambas delimitando o tema.

A figura 9, apresenta o gráfico com os principais autores, os que mais publicam sobre o assunto, é possível observar que apesar das principais publicações estarem nos Estados Unidos, os três principais autores são de origem chinesa, este fato guarda relação com o gráfico da figura 11, onde os países mais citados são os Estados Unidos com 1.086 e a China em segundo lugar com 1.010 citações no ambiente global, neste contexto o Brasil figura na 5ª posição com 500 citações.

A figura 10, apresenta a produção dos autores em função da lei de *Lotka*, de acordo com essa Lei, a produtividade dos autores segue uma distribuição logarítmica inversa, ou

seja, um pequeno número de autores produz uma quantidade significativa de publicações, enquanto a maioria dos autores produz um número relativamente baixo de publicações. Neste contexto a figura 9, corrobora com o observado na figura 10, em que um autor publicou nove artigos, três autores publicaram seis artigos, seis autores publicaram 5 artigos e o grande maioria publicou apenas um artigo.

A figura 12, apresenta em forma de nuvem de palavra os termos mais usados no conjunto dos artigos analisados, aparecendo com maior frequência o termo *classification* sendo mencionada 70 vezes, em segundo lugar aparece *prediction* sendo mencionada 41 vezes e terceiro lugar, temos o termo *model* sendo usada 40 vezes. Desta forma, os termos mencionados guardam relação com o tema.

Por fim, temos a figura 13, que apresenta em forma de gráfico as redes de autores, a relação entre eles e os grupos que trabalham isoladamente publicando seus artigos em muitos dos casos nos mesmo periódico sendo o caso do BISHOP, JC e YANG, GJ que publicam no periódico *Computers And Electronics In Agriculture*.

A análise bibliométrica apresentada no texto revela diversas informações sobre a produção científica relacionada a um determinado tema. A partir das figuras apresentadas, é possível observar a distribuição de produtividade dos autores, que segue a Lei de *Lotka*, e identificar os principais autores e países que publicam sobre o assunto. Além disso, a nuvem de palavras destaca os termos mais utilizados no conjunto dos artigos analisados, enquanto a rede de autores mostra a relação entre eles e os grupos que trabalham isoladamente publicando em periódicos comuns.

Portanto, destaca-se também a importância da análise bibliométrica para a avaliação da produtividade e impacto de autores e instituições em um determinado campo científico, bem como para a identificação de padrões de colaboração entre pesquisadores. Por fim, a análise bibliométrica apresentada no texto pode fornecer informações úteis para a tomada de decisões na pesquisa científica e para a elaboração de políticas públicas relacionadas à ciência e tecnologia.

## REFERÊNCIAS

BEUREN, Ilse M. Em busca de um delineamento de proposta para classificação dos periódicos internacionais de contabilidade para o qualis capes. **Revista Contabilidade & Finanças**, São Paulo, 15 agosto 2007. 46.

BIANCHI, José A. D. S. M. D. L. P. Cientometria: a métrica da ciência. **SciELO - Scientific Electronic Library Online**, São Paulo, Maio 2002. 5-10.

BOUSTANY, Joumana. La production des imprimés non-périodiques au Liban de 1733 a 1920: étude bibliométrique. **Tese (Doutorado em Sciences de l'Information et de la Communication) – Université Michel de Montaigne**, Bordeaux III, 1997.

BOYACK, K.W., K. R. A. B. K. Mapping the Backbone of Science. **Scientometrics**, Albuquerque, agosto 2005. 351-374.

ČATER, Ivan Z. T. Bibliometric methods in management and organization. **Organizational Research Methods**, London, julho 2015. 429-472.

FERENHOF E FERNANDES, Helio A.. R. F. F. Desmistificando a revisão de literatura como base para redação científica: método SSF. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, 11 agosto 2016. 550-563. Disponível em: <https://revista.acbsc.org.br/racb>. Acesso em: 23 fevereiro 2023.

LOTKA, A.J. The Frequency Distribution of Scientific Productivity. **Journal of the Washington Academy of Sciences**, Washington, 16, n. 12, 19 junho 1926. 317-323. Disponível em: <https://www.jstor.org/stable/24529203>. Acesso em: 02 março 2023.

MICHAEL, POLANYI. J. Z. A. S. F. THE REPUBLIC OF SCIENCE: ITS POLITICAL AND ECONOMIC THEORY Minerva. **Minerva**, Spring Street, janeiro 2000. 54-73.

PRITCHARD, A. Statistical bibliography or bibliometrics? **Journal of Documentation**, London, 10 January 1969. p. 348-349.

SOARES, Patrícia B. et al. Análise bibliométrica da produção científica brasileira sobre Tecnologia de Construção e Edificações na base de dados Web of Science. **Ambiente Construído**, Porto Alegre, 10 março 2016. 175-185.