InterSciencePlace

International Scientific Journal – ISSN: 1679-9844
Nº 3, volume 18, article nº 01, July/September 2023
D.O.I: http://dx.doi.org/10.6020/1679-9844/v18n3a1
Accepted: 10/04/2022 Published: 20/09/2023

BIBLIOMETRIC STUDY: MACHINE LEARNING APPLIED IN AGRICULTURE

Gizele Ferreira da Silva[1]
[1]Federal University of Rondônia Foundation. E-mail: gizele100@hotmail.com

David Lopes Maciel[2]
[2]Professional Master in Master of Science in Emergent Technologies in Education. MUST UNIVERSITY, MUST, EUA. E-mail: maciel.1000@hotmail.com

Carlos Alberto Paraguassu-Chaves[3]
[3]PhD in Health Sciences, University of Brasília - UnB, Brazil and PhD in Science - University of Havana, Cuba and Post-Doctor in Health Sciences, UnB and Degli Studi D'Aquila University, IT, Professor at the University Institute of Rio de Janeiro, IURJ, Brazil

Fabrício Moraes de Almeida[4]
[4]PhD in Physics (UFC), with post-doctorate in Scientific Regional Development (DCR/CNPq). Specialist in Software Engineering (FUNIP). Researcher of the Doctoral and Master Program in Regional Development and Environment (PGDRA/UFRO). E-mail: dr.fabriciomoraes001@gmail.com

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 [9]. A Bibliometria faz uso de métodos matemáticos afim de descrever e quantificar estudos relacionados a uma temática científica [10]. Neste contexto, o objetivo deste paper é quantificar as publicações na área de *Machine Learning* 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 analises 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: Estudos bibliométricos. Machine learning. Agricultura.

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 [9]. Bibliometrics makes use of mathematical methods in order to describe and quantify studies related to a scientific theme [10]. In this context, the objective of this paper is to quantify the publications in the area of Machine Learning 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: Bibliometric studies. Machine Learning. Agriculture.

1. INTRODUCTION

This paper deals with the analysis of the world scientific production on the application of artificial intelligence in agriculture through bibliometric indicators. The theme was chosen because it is in evidence and presents great relevance for the productive environment of agribusiness, configuring itself in a strategic area in food production worldwide, in this context the improvement and expansion of food production is of general interest, meets objective two of the 2030 agenda.

To carry out the study, bibliometrics was used, which is a method of quantitative analysis for scientific research. The data elaborated through bibliometric studies estimate the contribution of scientific knowledge to the academic environment and society, deriving from the databases the publications, and from the set of them the indicators used in the analyses and discursions [1].

Also according to scientometrics, which is defined as the study of the measurement and quantification of scientific progress, and research is based on bibliometric indicators, scientific mapping uses bibliometric methods to examine disciplines, fields, specialties, authors and research networks, and seeks to identify how they relate to each other [2].

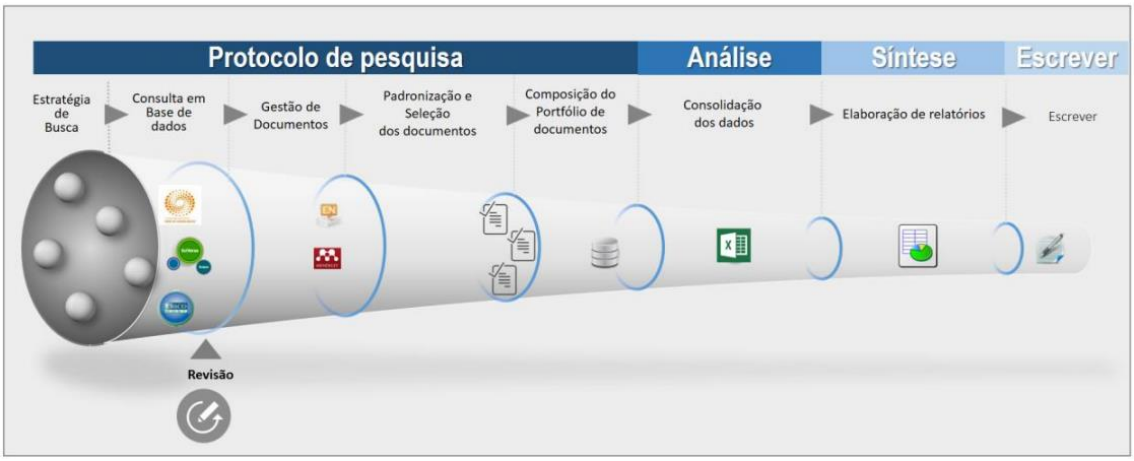
In this way, the method produces maps and spatial and temporal representations of the data obtained in the databases. Corroborates this definition, Boyack and Klavans when they state that "[...] Scientific mapping is a combination of classification and visualization" [3]. The goal is to create a representation of the structure of the research area, partitioning elements (documents, authors, journals, words) into different groups. The visualization is then used to create a visual representation of the classification that emerges [4].

The softwares used in this work were R / RStudio / Bibliometrix / Biblioshine in conjunction with MS Excel, the databases used were Web of Science and Scopus, for presenting the largest number of indexing of journals. The research was systematized through the SSF - Systematic-Search Flow method, which aims to systematize the search process in order to ensure repeatability and replicability, avoiding biased biases on the part of the researcher [5].

2. METHODOLOGY

In this work, the operationalization of the research took place through the application of the SSF method, which consists of 4 phases and 8 activities, as presented in Fig. 1, the method was developed according to the authors, in order to systematize the process of search or searches to the scientific database. Thus, it serves both for the systematic review and for the integrative review, depending only on the definition of the strategy in its use [5].

Figure 1 - Representation of the Systematic Search Flow Method.



Source: Ferenhof and Fernandes [5].

It is worth mentioning that at the time of the use of the R® software, in phase 1 (Research Protocol) activities 3, 4 and 5 were fully automated with the use of the RStudio / Bibliometrix Libraries. In phase 2 (Analysis) the activity (data consolidation) was also automated with the same libraries of R®, however a database file was generated in MS Excel®, containing all the articles searched and unifying the data obtained in the Web of Science and Scopus database, this served to treat again the data from Bibliometrix. Phase 3 (Synthesis) followed the same procedure of automating the process with the use of the R® library/Biblioshine applied to the Database generated in the previous phase.

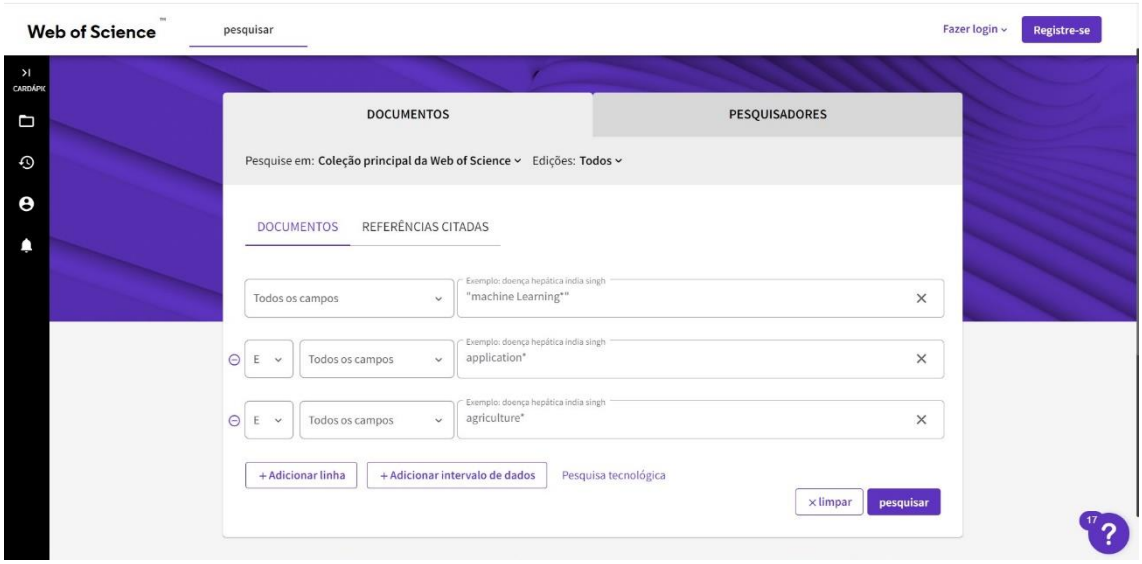
In phase 4, as recommended by the authors Ferenhof and Fernandes, the consolidation of the results was carried out through scientific writing, therefore, the objective of the work was rescued, compared with the result of the analysis and synthesis, in this phase, the knowledge matrix recommended by the authors was eliminated, since the data were generated within the Biblioshine, as well as the reports that supported the writing of the results, ending activity 8, proposed by the authors.

2.1 – Search Execution

Research protocol:

- Activity 1 - search strategy making use of the logical operator (AND, *, " "), and the search Query (articles, documents in English, articles indexed in the period between 2015 and 2022, and application of filters in the categories of the Web of Science);
- Activity 2 - parameterization and application of the search query in the selected base according to figures 2, the terms used were (machine learning, applica-tion and agriculture).

Figure 2 - Parameterization of the search in the Web of Science Database.



Source: authors, 2023.

The research was carried out in the Web of Science database, through the CAPES journal portal – CAFe.

3. RESULTS

From the initial search, the database was generated containing the information of articles indexed in the Web of Science database with 476 articles downloaded in the extension "BiTex" as shown in Figure 3 – Print of the Search result. After downloading, the database was exported to Bibliometrix R, where it underwent treatment generating an MS Excel® file of the base. Again this file went through a reanalysis and treatment where the criterion was to keep only the articles that had DOI, being also eliminated the duplicates, in this way, 475 articles remained, being eliminated 1 article for not having DOI. Subsequently, these were exported to the R – Biblioshine interface, where the information and the various graphs were generated for further analysis of the results.

Figure 3 – Print of the Search result.

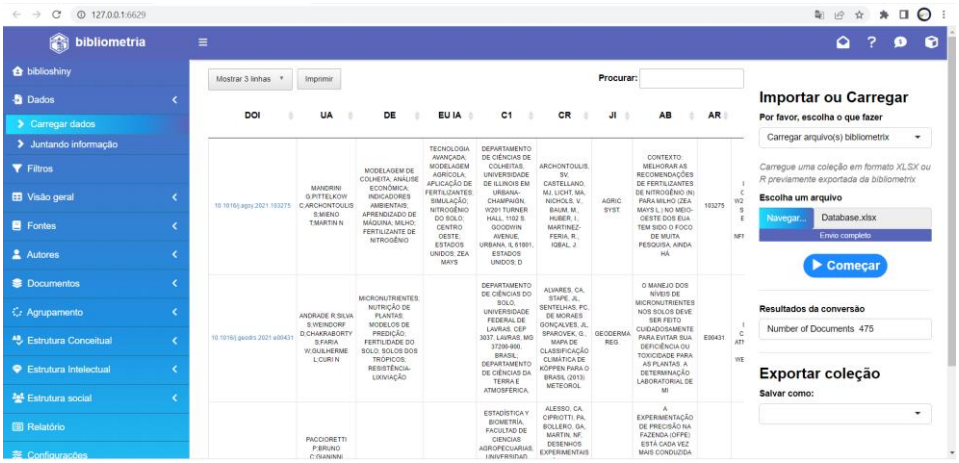
Source: authors, 2023.

Figure 3 shows the data obtained in the Web of Science database, containing the 476 articles, the figure also shows the configuration of the search query and the filters applied.

3.1 – Results of processing in Biblioshine.

After data processing in the R / Bibliometrix library, the following results were obtained in figures 4, 5, 6, 7, 8, 9, 10, 11, 12.1, and 12.2.

Figure 4 – Biblioshine import interface.



Source: authors, 2023.

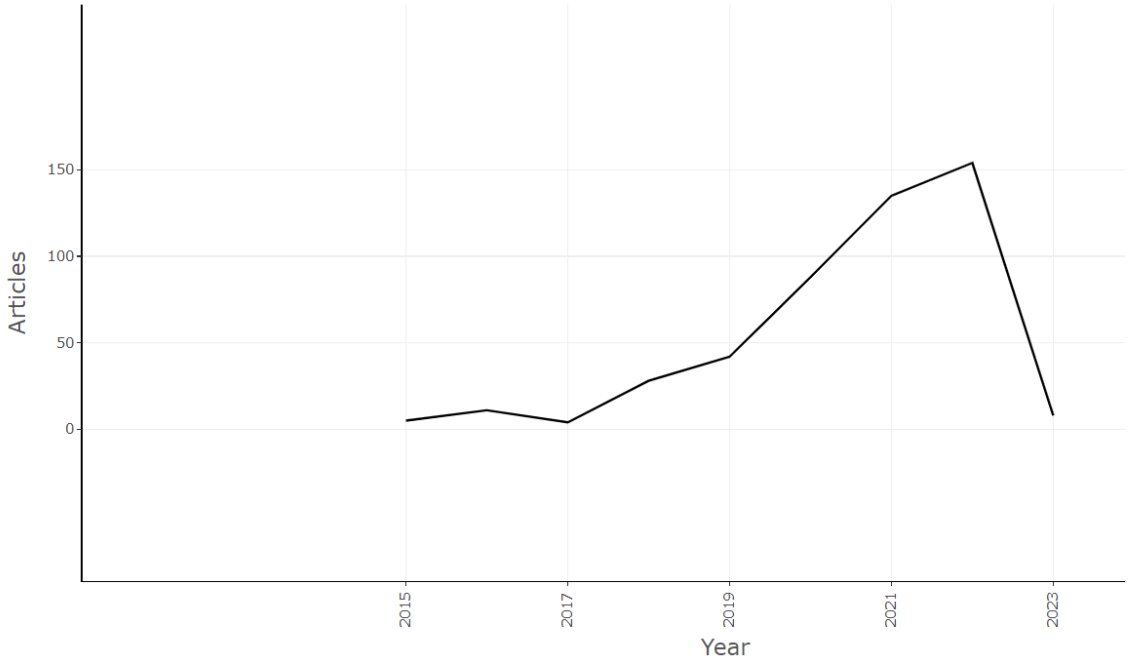
In the data import interface, it is possible to verify the data that were obtained and correlate them with the base file, such as time interval of the selected documents, number of authors, references and documents, types of documents, among others.

Figure 5 – Main Information.

Source: authors, 2023.

Figure 5 presents in numbers the information in general, and a summary of the information obtained.

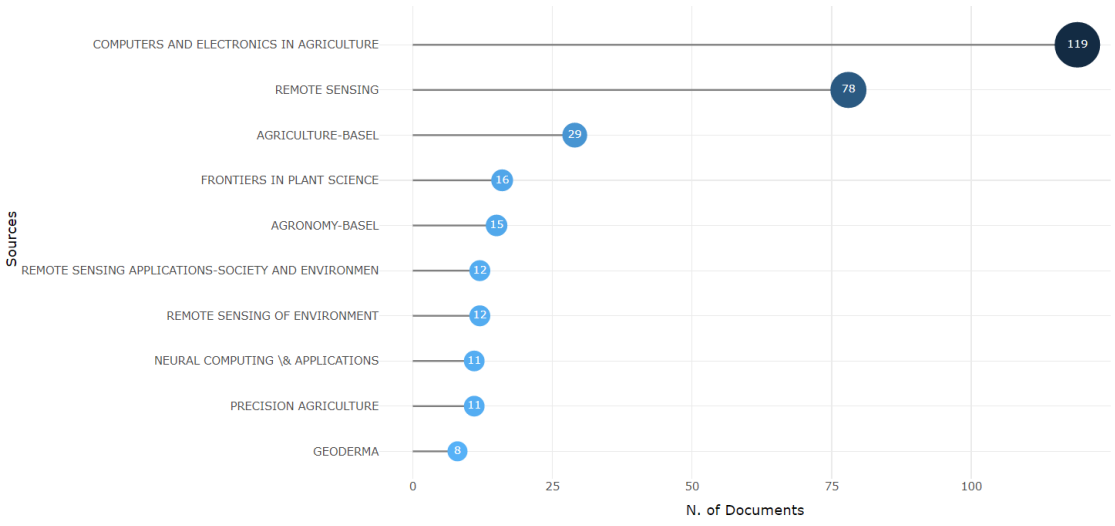
Figure 6 – Annual Scientific Production.



Source: authors, 2023.

Figure 6 shows the information of the annual scientific production.

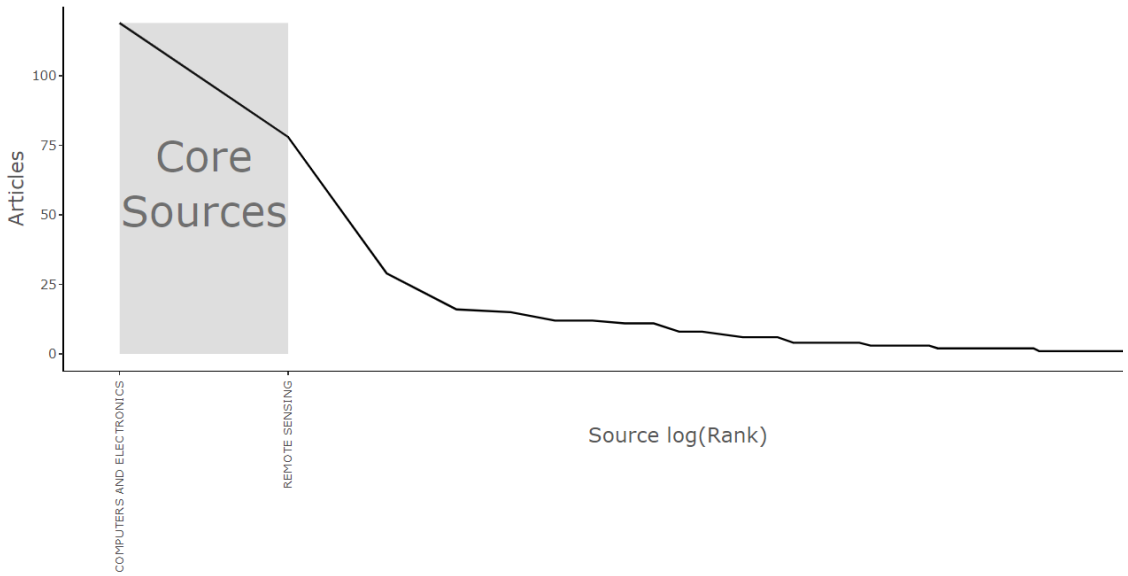
Figure 7 – Most Relevant Sources.



Source: authors, 2023.

Figure 7, on the other hand, shows the graph with the indication of the most relevant sources worldwide.

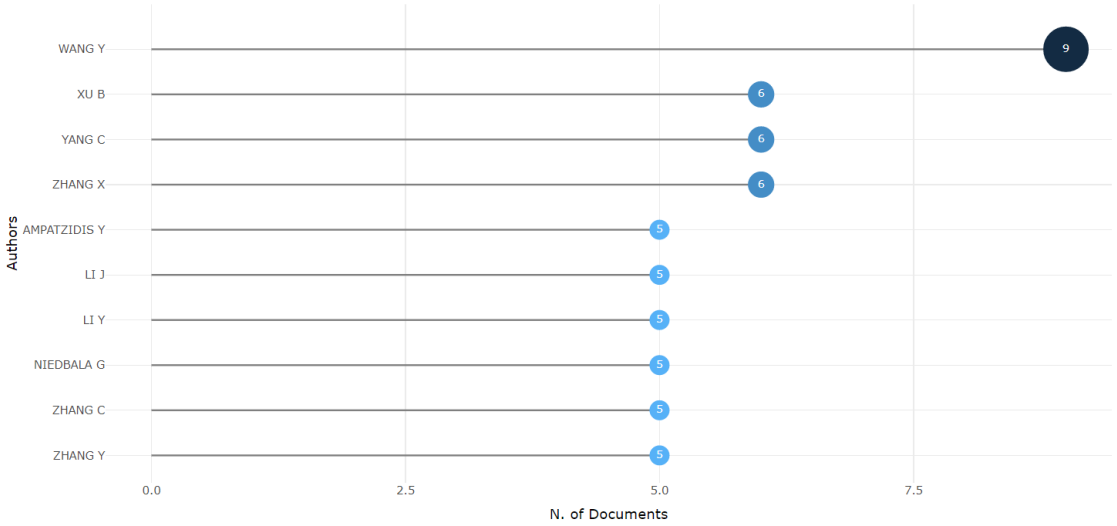
Figure 8 – Core Sources by Bradford's Law.



Source: authors, 2023.

In Figure 8, the graph shows the publications taking into account the main sources categorized by Bradford's Law. This law provides for the decreasing ordering of the productivity of articles on a given subject in scientific journals, allowing the establishment of exponentially divided groupings.

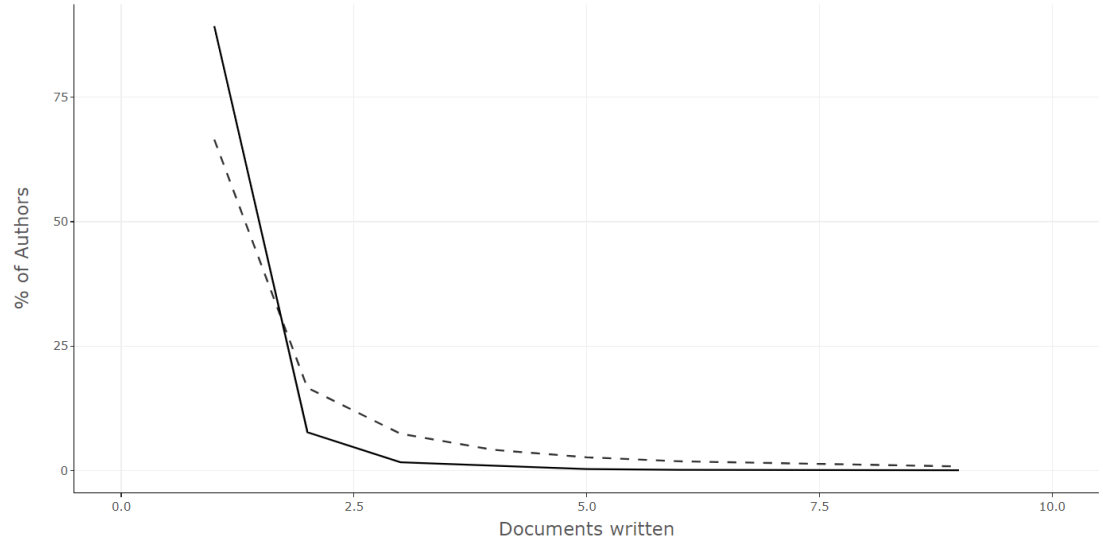
Figure 9 – Most Relevant Authors.



Source: authors, 2023.

Figure 9 shows the graph with the main authors with greater relevance in the international scenario.

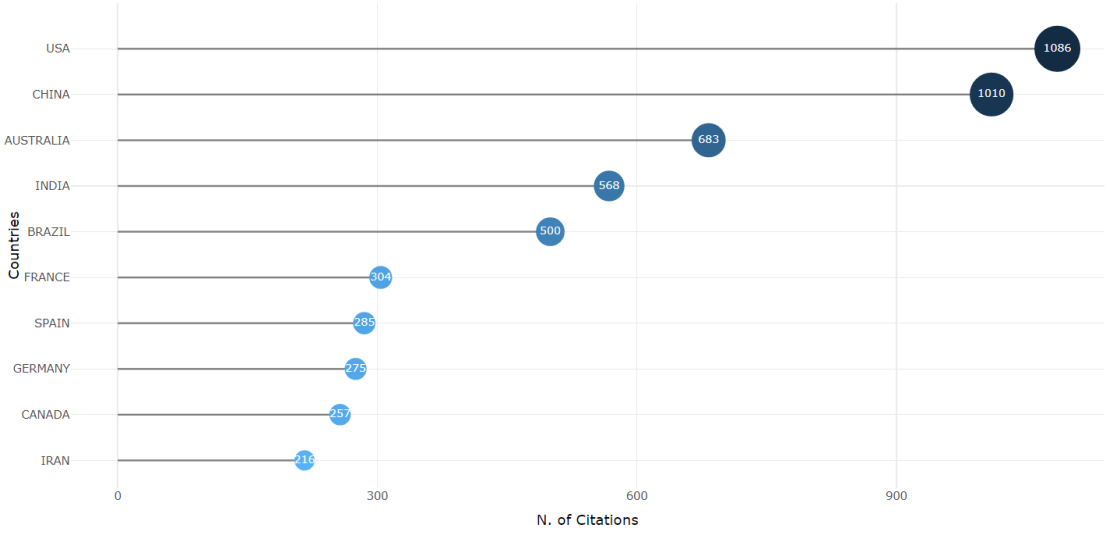
Figure 10 – Author Productivity through Lotka's Law.



Source: authors, 2023.

Figure 10, brings us the productivity of authors analyzed through Lotka's Law, this law describes the frequency of publication of authors in any field of knowledge, based on the inverse square, in which the number of authors who publish a certain number of articles is a fixed proportion to the number of authors who publish a single article.

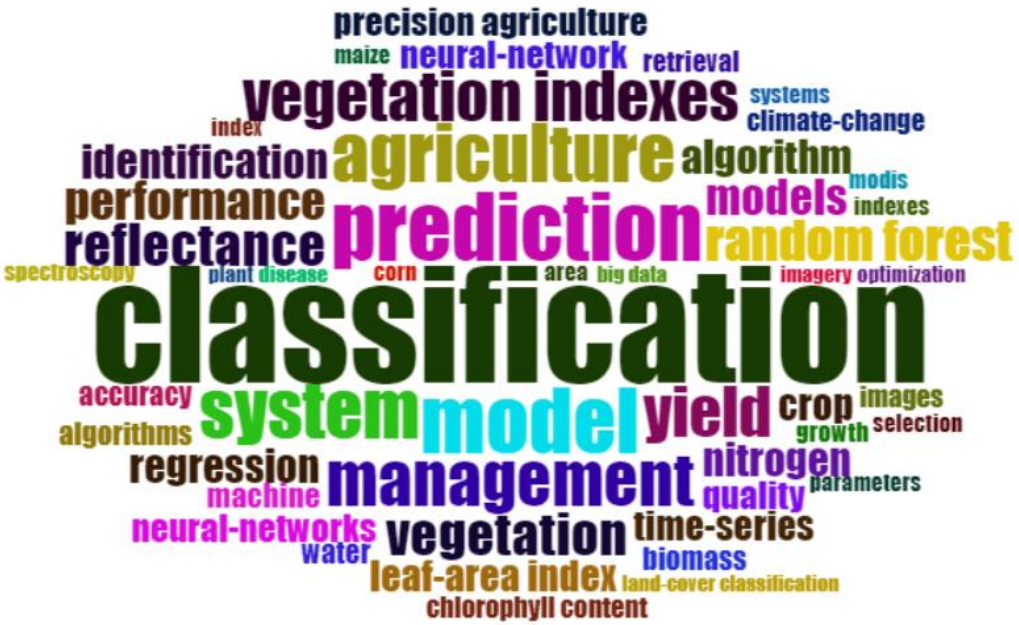
Figure 11 – Most Cited Countries.



Source: authors, 2023.

Figure 11 shows the graph of the most cited countries in scientific publications and the frequency of citations.

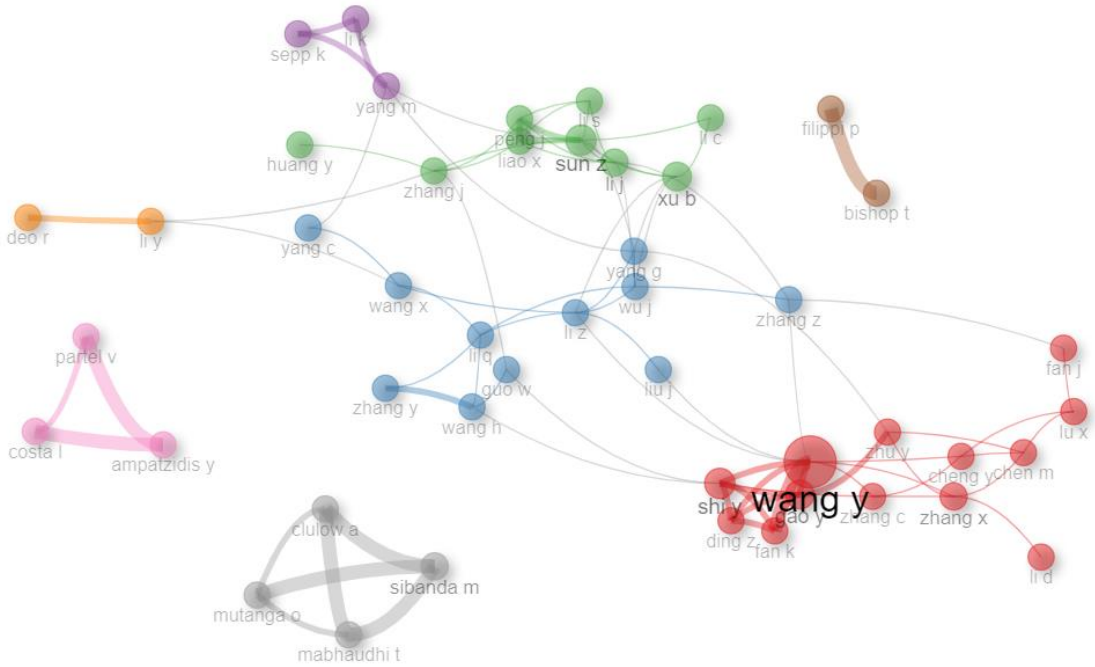
Figure 12 – WordCloud.



Source: authors, 2023.

Figure 12.1 presents in the form of a word cloud, the most used terms in publications, as well as the frequency of their occurrence. The larger the word size, the higher the frequency of use, so the smaller the word size, the lower its frequency.

Figure 12.2 – Collaboration Network.



Source: authors, 2023.

Finally, Figure 12.2 graphically presents the internal networks of authors (without connection with other networks), allowing us to show which authors and networks have published the most on the subject addressed.

4. CONCLUSION

The use of bibliometric indicators to study the research activities of one or more fields of knowledge has become increasingly common in the evaluation of scientific production. These indicators may include information on the number of articles published, the frequency of citations received, the impact of the journals in which they were published, among other data.

Bibliometrics allows, therefore, a quantitative analysis of scientific production, which can be useful both for the comparison between different areas of knowledge and for the individual evaluation of researchers and institutions. However, it is important to remember that bibliometrics should not be used as the only source of evaluation, but should be complemented by other qualitative and subjective methodologies.

From a cognitive point of view, new knowledge only acquires its value when it is disseminated within the scientific community. This is because knowledge production is a social process, and it is only through interaction between researchers that new ideas and discoveries can be evaluated, tested, and integrated into existing knowledge [6].

This leads to the affirmation that the reliability or reliability of the results of bibliometric studies will depend substantially on their correct application, taking into account their advantages, but also the limitations and necessary conditions of their use [4].

In this context, the bibliometric study carried out, using the SSF method of Ferenhof and Fernandes [5] and adapted by the authors of the article, made it possible to obtain, treat and analyze the data, as well as the generation of the results that we are now analyzing. Based on the information obtained and presented in Figure 6, the theme is current and presents a constant growth pattern, which can be observed in Figure 5, where the growth rate is 6.05% per year, with an average age of 2.39 years in the sets of documents.

Figure 7, presents the graph with the indication of the sources with greater global relevance, it is possible to observe that when it comes to the theme, multidisciplinary journals encompass most of the articles, being the case of the journal Computers and Electronics in Agriculture - Copyright © 2023 Elsevier B.V. All, which provides international coverage of advances in the development and application of computer hardware, software, electronic instrumentation and control systems to solve problems in agriculture, including agronomy and horticulture, stands out with 119 publications involving the proposed theme.

Figure 8 shows the graph of the publications taking into account Bradford's law. Bradford's Law suggests that as the first articles on a new subject are written, they are subjected to a small selection, by appropriate journals, and if accepted, these journals attract more and more articles in the course of the development of the area/subject. At the same time, other journals publish their first articles on the subject [7], which can be confirmed in the chart with Computers and Electronics in Agriculture, followed by the journal REMOTE SENSING- MODIS (Moderate Resolution Imaging Spectroradiometer), both delimiting the theme.

Figure 9 shows the graph with the main authors, the ones who publish the most on the subject. It is possible to observe that although the main publications are in the United States, the three main authors are of Chinese origin, this fact is related to the graph in Figure 11, where the most cited countries are the United States with 1,086 and China in second

place with 1,010 citations in the global environment, in this context Brazil is in the 5th position with 500 citations.

Figure 10 shows the authors' production according to Lotka's law [8]. According to this Law, the productivity of authors follows an inverse logarithmic distribution, that is, a small number of authors produces a significant number of publications, while most authors produce a relatively low number of publications. In this context, Figure 9 corroborates what was observed in Figure 10, in which one author published nine articles, three authors published six articles, six authors published 5 articles and the vast majority published only one article.

Figure 12.1 presents in word cloud form the most used terms in the set of articles analyzed, appearing more frequently the term classification being mentioned 70 times, in second place appears prediction being mentioned 41 times and third place, we have the term model being used 40 times. In this way, the terms mentioned are related to the theme.

Finally, we have Figure 12.2, which presents in graphic form the networks of authors, the relationship between them and the groups that work in isolation publishing their articles in many of the cases in the same journal being the case of BISHOP, JC and YANG, GJ that publish in the journal Computers And Electronics In Agriculture.

The bibliometric analysis presented in the text reveals several information about the scientific production related to a given theme. From the figures presented, it is possible to observe the distribution of productivity of the authors, which follows Lotka's Law, and identify the main authors and countries that publish on the subject. In addition, the word cloud highlights the most used terms in the set of articles analyzed, while the network of authors shows the relationship between them and the groups that work in isolation publishing in common journals. The author Boustany [9] highlights the indicators of scientific production and Pritchard [10] when making use of mathematical methods to describe and quantify scientific studies, corroborate with the bibliometric analyses.

The importance of bibliometric analysis for the evaluation of the productivity and impact of authors and institutions in a given scientific field is also highlighted, as well as for the identification of patterns of collaboration between researchers. Finally, the bibliometric analysis presented in the text can provide useful information for decision-making in scientific research and for the elaboration of public policies related to science and technology.

REFERENCES

[1] 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.

[2] 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.

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

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

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

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