Bibliometric analysis of publications on the use of artificial intelligence in digital marketing

Análisis bibliométrico sobre publicaciones del uso de la inteligencia artificial en el marketing digital

GUTIERREZ-BELTRAN, Brenda Yadira†, GOMEZ-BARBA, Leopoldo* and PRECIADO-ORTIZ, Claudia Leticia

Universidad de Guadalajara, Doctorado en Tecnologías de Información, Doctorado en Ciencias de la Administración, CUCEA, México.

ID 1st Co-author: *Brenda Yadira, Gutierrez-Beltrán* / **ORC ID:** 0009-0002-0642-4608

ID 1st Co-author: *Leopoldo, Gómez-Barba* / **ORC ID:** 0000-0003-0608-0255

ID 2nd Co-author: *Claudia Leticia, Preciado-Ortíz* / **ORC ID:** 0000-0003-2391-2734

Abstract

This study analyzes the evolution of research on machine learning in marketing from 1991 to November 16, 2022. The search was carried out in several databases and the terms "machine learning" and "marketing" were used. 2,358 articles were found, from which important indicators were obtained: 2021 has been the year with the highest publication activity with 234 articles. The most productive authors are from various fields, including computer science, marketing, and data analysis. The most cited article focuses on the application of deep learning in the fashion industry. The results suggest that research on aprendizaje automático in marketing is growing rapidly and has the potential to transform marketing practices. The need for more research is highlighted to assess the quality of the research and its impact on the scientific community. Bibliometric analysis is a tool that allows the researcher to carry out bibliographic searches effectively, since it generates an overview of the trends in research on the subject

Machine learning, Marketing, Bibliometric Analysis

Resumen

Este estudio analiza la evolución de la investigación sobre el aprendizaje automático en marketing desde 1991 hasta el 16 de noviembre de 2022. La búsqueda se realizó en varias bases de datos y se utilizaron los términos "aprendizaje automático" y "marketing". Se encontraron 2,358 artículos, de los cuáles se obtuvieron importantes indicadores: el año 2021 ha sido el de mayor actividad de publicación con 234 artículos. Los autores más productivos son de diversos campos, incluyendo ciencias de la computación, marketing y análisis de datos. El artículo más citado se centra en la aplicación del aprendizaje profundo en la industria de la moda. Los resultados sugieren que la investigación sobre el aprendizaje automático en marketing está creciendo rápidamente y tiene el potencial de transformar las prácticas de marketing. Se destaca la necesidad de más investigación para evaluar la calidad de la investigación y su impacto en la comunidad científica. El análisis bibliométrico es una herramienta que permite al investigador realizar búsquedas bibliográficas de manera eficaz, ya que genera un panorama general sobre las tendencias en la investigación sobre el tema.

Aprendizaje automático, Marketing, Análisis bibliométrico

---

* Author's Correspondence (E-mail: brenda.gutierrez4390@alumnos.udg.mx)

† Researcher contributing as first author.

Introduction

One of the most significant goals of research is the dissemination of results through documents (scientific publications), where observations and achievements of work are compiled to be shared with the wider community, thereby generating a positive impact and fostering growth or enhancement within society. Furthermore, the sheer volume of publications stemming from the eager pursuit of knowledge drives us into an escalating spiral of information that is more accessible today than ever before through electronic means, resulting in the quest for research output with impact and quality becoming exhaustive and possibly unattainable in its entirety to encompass the breadth of the subject of interest. Consequently, it becomes advantageous to employ mechanisms for analyzing the quantity of products generated as a consequence of the research process, in order to assess the quality of the knowledge generation process and its impact within the scientific realm (Christian Federico Rueda-Clausen Gómez et al., 2005).

The acquisition of bibliometric indicators serves as a means to measure growth in scientific production, while also enabling the analysis of the research's impact within the universe of study. This analysis relies on quantitative values derived from production, its interrelation with other fields, areas of heightened interest, and more, all of which are derived from scientific publications.

Bibliometrics stands as a subdiscipline of scientometrics, providing insights into the outcomes of the research process, including the volume of output, its temporal evolution, the attained visibility, and the structure of the research (Tatiana Alejandra Escorcia Otalora, 2008). On the other hand, bibliometric analysis serves as the means to obtain and analyze indicators such as the number of articles, impact factor, productivity index, immediacy, collaboration index, among others (Christian Federico Rueda-Clausen Gómez et al., 2005).

In the present study, bibliometric analysis has been employed to gather information regarding machine learning in the context of marketing and to identify the relationship between these two fields of study.

This analysis will be utilized to infer areas of opportunity and construct a model focused on the application of machine learning to address marketing challenges, as the topic of Innovation in Digital Marketing has been a prominent research trend over the last 5 years (2018-2022) (Muhammad Rosyihuddin & Sudarmiatin, 2023).

Data and methods

The data for this study were sourced from Scopus as of November 16, 2022. Scopus was chosen as the search engine due to its status as one of the largest databases of peer-reviewed scientific literature and citations. The search terms "machine learning" and "marketing" were utilized as the search topic. This topic-based search indicates that the terms "machine learning" and "marketing" are contained within the title, abstract, and/or keywords of the publications.

Search results

The time period considered ranged from 1991 to November 16, 2022, serving as the cutoff date. It is worth noting that if the same search were conducted on a different date, the results might diverge due to Scopus being a dynamically updated database. In total, 2358 publications related to machine learning in marketing were identified. The search encompassed all types of publications. Examining the document types, the majority consisted of conference papers (n = 1047) and articles (n = 1038), with the remainder comprising book chapters, books, and reviews, among others. Each publication in Scopus contains extensive details, including the publication year, authors, author affiliations, title, abstract, source journal, subject categories, and references. The data from these 2358 publications stored in Scopus were exported to Excel.

A spreadsheet and statistical software were employed to analyze the following indicators: (1) article count, (2) productivity index, (3) impact factor, (4) immediacy, (5) median age of citations, (6) collaboration index, (7) source data, and (8) keywords.

Results and discussions

Article count

In this indicator, the number of articles focusing on machine learning in marketing was tallied from 1991 to 2022, providing insights into the annual publication volume (Christian Federico Rueda-Clausen Gómez et al., 2005). As depicted in Figure 1, the count of publications concerning machine learning in marketing has steadily increased since 2005. There was only one publication on machine learning in marketing in 1991. Up until 2004, publications on the topic remained limited (no more than 10 publications per year). Starting from 2005, a consistent growth in the number of publications was observed each year, with exceptions in 2006, 2010, and 2013, when a decline was evident. The peak of publications was reached in 2021 (n = 506), followed by a subsequent decline in the trend (n = 399 in 2022). Examining the cumulative number of publications in Figure 2, it can be affirmed that the significance of research on machine learning in marketing has experienced a substantial rise. It took approximately twelve years (from 1991 to 2005) to accumulate a total of 56 publications within the field of machine learning in marketing. In the ensuing four years (from 2006 to 2022), this number surged to a cumulative total of 2358 publications.
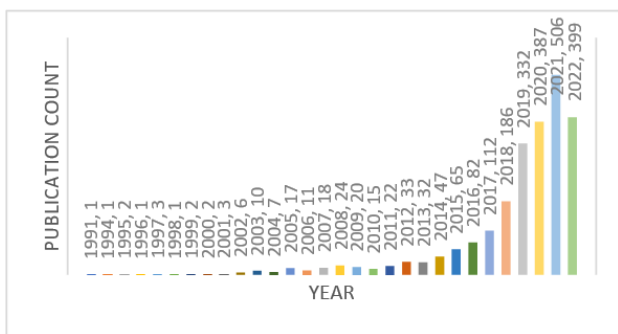
Graphic 1 Indicator: article count, crafted in spreadsheet
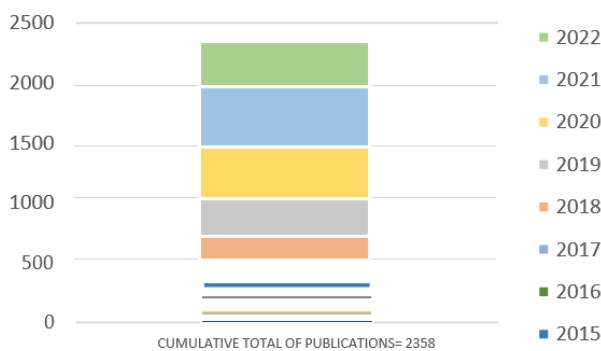
Graphic 2 Cumulative publications from 1991 to 2022, created in spreadsheet

Productivity index

This indicator is widely used in productivity assessment and is measured by the number of publications per researcher, institution, or group. It is governed by Lotka's law, as it serves as an instrument for describing thematic areas and is defined as the decimal logarithm of the number of articles produced (Tatiana Alejandra Escorcia Otalora, 2008). Personal factors, certain environmental and situational aspects, as well as behavioral factors, also influence the number of publications by individuals. The most influential factor is funding (Wahid et al., 2022). The 2358 publications were authored by a total of 159 different authors. The majority of authors, 51.57% (n = 82/159), are credited with only two publications on the subject of machine learning in marketing, while 73.58% (n = 117/159) are credited with at least three publications. These authors are categorized as intermediate producers. Additionally, 0.628% (n = 1/159) are credited with over 10 publications and are considered significant producers (Tatiana Alejandra Escorcia Otalora, 2008). Table 1 presents the top 3 most productive authors in the field of machine learning in marketing. This index is based on the total number of authors and publications and not on author order. Kar A.K. is the most productive author in the field of machine learning in marketing, with eleven publications, followed by Lessmann, S., with nine, and Goto, M., and Salas-Rueda, R.A., each with six publications. The top 3 authors have an average of 12.94 publications.

Author name	Number of publications	Productivity percentage
Kar, A.K.	11	0.47%
Lessmann, S.	9	0.38%
Goto, M.	6	0.25%
Salas-Rueda, R.A.	6	0.25%

Table 1 Most productive authors utilizing machine learning in marketing
Source: Own elaboration

Impact factor

While the impact factor of journals remains the most debated bibliometric indicator and has exerted a significant influence on the scientific domain (Larivière & Sugimoto, 2019), it allows for the assessment and comparison of a journal's significance relative to others within its field (Christian Federico Rueda- Clausen Gómez et al., 2005).

GUTIERREZ-BELTRAN, Brenda Yadira, GOMEZ-BARBA, Leopoldo and PRECIADO-ORTIZ, Claudia Leticia. Bibliometric analysis of publications on the use of artificial intelligence in digital marketing. Journal-Business Administration-Marketing; Accounting. 2023

It represents the quotient between the number of citations to articles from the given journal and the total number of articles published within the past two years. In total, out of the 2358 publications, 953 were published in journals, indicating a diverse range of research topics and the multidisciplinary nature of machine learning research in marketing. Among these 953 publications, 694 journals (72.82%) contributed only one publication, while 131 journals (13.74%) published two articles solely on the topic of machine learning in marketing. Furthermore, 101 journals published between 3 and 9 articles (10.61%), and 27 journals published ten or more articles on the subject (02.83%). Table 2 provides insights into the 10 most active journals in the field of machine learning research in marketing. These ten journals, accounting for only 1.04% of all journals publishing on machine learning in marketing, have contributed over one-fifth of all machine learning in marketing publications (22.36%; n= 454/2030). The journal "Journal of the Academy of Marketing Science" stood out with the highest impact factor regarding the 2021 published article "A strategic framework for artificial intelligence in marketing," garnering 113 citations during 2021 and 2022 out of a total of 905, yielding an IF = 8.0088, signifying an average of 8 publications per year.

RANK	JOURNAL TITLE	NUMBER OF PUBLICATIONS
1	Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics	72
2	Advances In Intelligent Systems And Computing	63
3	ACM International Conference Proceeding Series	62
4	Lecture Notes In Networks And Systems	36
5	Communications In Computer And Information Science	32
5	Expert Systems With Applications	32
6	Ceur Workshop Proceedings	28
7	IEEE Access	24
7	Journal Of Physics Conference Series	24
8	Journal Of Business Research	23
8	Smart Innovation Systems And Technologies	23
9	Lecture Notes In Electrical Engineering	18
10	International Journal Of Advanced Computer Science And Applications	17

Table 2 Top 10 Journals with Most Publications on Machine Learning in Marketing
Source: Own elaboration

Immediacy

Immediacy is defined as the ratio of citations received by works published during a given year, divided by the total number of articles published in the same journal and year (Tatiana Alejandra Escorcia Otalora, 2008).

In the case of the year 2021 (see Figure 1), during which a higher number of publications on machine learning in marketing occurred, the "Journal of the Academy of Marketing Science" published 1 article, which received 113 citations. Additionally, the journal accumulated a total of 2216 citations across 65 publications within that same year, resulting in an immediacy index of 34.09 and an equivalent impact factor of 13.87, as reported on the Scimago Journal Rank website.

Median citations age

This metric aids in evaluating the range of the age of journal articles that have been cited, and it can be used to make decisions regarding archiving and preservation (Eduardo Luis De Vito, 2006). Publications concerning machine learning in marketing have an average age of 4.43 years. This indicator represents the average age across each publication. The researcher employed the Scopus database to analyze the number of times machine learning in marketing publications were cited.

In total, the 2358 publications were cited 30,517 times as references in other publications. The average citation per publication is 12.94. Of the total, 36.51% (n = 861/2358) received zero citations up until the cutoff date of November 16, 2022.

Among the 2358 machine learning in marketing publications, 57.97% (n = 1367/2358) received between 1 and 49 citations, 03.30% (n = 78/2358) received between 50 and 99 citations, and 02.20% (n= 52/2358) received 100 citations or more.

Table 3 presents the ten most cited publications. The most cited article is "Machine Learning: Trends, Perspectives, and Prospects" by Jordan M.I., Mitchell T.M. This article has been cited 2722 times since its publication in 2015 (up to November 16, 2022). It has also been the publication with the highest average citations per year, with an average of 544.4 citations annually.

Rank	Title	Citations
1	Machine Learning: Trends, perspectives, and prospects	722
2	A survey of text classification algorithms	1087
3	Ad click prediction: A view from the trenches	529
4	Computer-based personality judgments are more accurate than those made by humans	513
5	Field-aware factorization machines for CTR prediction	415
6	ls on travel search engines by mining user-generated and crowdsourced content	367
7	Classification of sentiment reviews using n-gram aprendizaje automático approach	356
8	Practical lessons from predicting clicks on ads at Facebook	335
9	Advertising content and consumer engagement on social media: Evidence from Facebook	321
10	The use of data mining and neural networks for forecasting stock market returns	292

Table 3 Top 10 Most Cited Publications
Source: Own elaboration

Collaboration Index

Studies on scientific collaboration hold a significant presence in bibliometric research and are commonly used as a measure of scientific collaboration among multiple authors, allowing for the determination of research group sizes (Tatiana Alexandra Escorcia Otalora, 2008). The average number of authors per publication was 3.4. Among the publications, 12.46% (n = 294/2358) were single-authored, 23.15% (n = 546/2358) had two authors, 24.59% (n = 580/2358) had three authors, and 39.77% (n = 938/2358) had four or more authors (with a maximum of 27 authors). Multi-author publications in the field of machine learning in marketing account for approximately 87.54% of the total. The collaboration patterns of authors publishing on machine learning in marketing were analyzed using Scopus. Authors in the collaboration network published at least two articles on the topic.

The author collaboration network is represented in Figure 1. Collaboration groups are denoted by colors, with the size of circles indicating the number of publications, and lines between two authors depicting their cooperation. Representing co- authorship collaboration, nine main author groups can be distinguished. The primary ones are Chen, Wang, and Zhang.
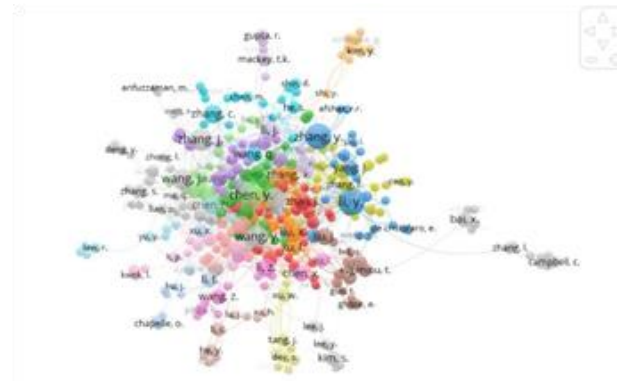

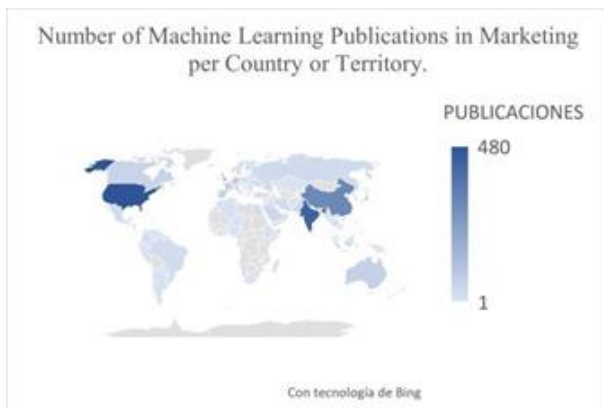**Figure 1** Collaboration Relationship Among Authors, Created in VOSviewer

Origin Data

Publications on Machine Learning in Marketing originate from 102 different countries or territories. Of these, 35.29% are located in Asia, 34.31% in Europe, 17.65% in the Americas, 10.78% in Africa, and 01.96% in Oceania. Figure 3 displays the number of publications on machine learning per country. Fifty- one countries (50%) produced 10 or more publications, while another fifty-one countries or territories (50%) have produced between 1 and 9 publications. The United States produced the highest number of publications (n = 480), followed by India (n = 434), and China (n = 305). Table 4 shows the top 10 most productive countries and territories in machine learning research in marketing. Economic development appears to contribute to scientific and academic investment, as all seven major industrialized countries in the world (G7: USA, Japan, Italy, Germany, UK, Canada, and France) ranked in the top 10 most productive countries publishing on machine learning in marketing. The dominance pattern of the G7 has been observed across most scientific fields, reflecting the high economic activity and level of these countries.(Y. Liu & Yang, 2022) . The total count of countries and territories from the 2975 publications with country or territory information was 2857 (an author may be affiliated with more than one country or territory, or a publication may be written by authors from different countries or territories). Expanding on the geographical information, an unequal distribution can also be observed across continents when examining the data: Asia accounted for 47.71% of the publications (n = 1363/2857), Europe for 25.06% (n = 716/2857), the Americas for 22.05% (n = 630/2857), Africa for 02.70% (n = 77/2857), and Oceania for 02.49% (n = 71/2857).

The cooperation network (i.e., co-authorship) among countries and territories publishing on machine learning in marketing was analyzed. Countries and territories in the network published at least ten publications on the subject. Countries or territories not connected with others in the network are not included. The outcome of the cooperation network among countries from each continent is presented in Figure 4. Colors represent countries from each continent that had publications on Machine Learning in Marketing. Funding, collaboration, and academic qualifications can influence a country's publication growth. The number of universities and ISI-indexed journals also contribute to the publication output of countries. (Wahid et al., 2022).

Graphic 3 Number of Machine Learning in Marketing Publications by Country or Territory, created using spreadsheet software

Rank	Country	Publications
1	United States	480
2	India	434
3	China	305
4	United Kingdom	111
5	Germany	78
6	Italy	70
7	Japan	64
7	Spain	64
8	Australia	61
9	Taiwan	57
9	Turkey	57
10	Canada	53

Table 4 Top 10 Countries with the Most Publications on Machine Learning in Marketing
Source: Own elaboration

Graphic 4 Continent-wise Country Grouping with Most Publications on Machine Learning in Marketing, created using spreadsheet software

Institutions

At the institutional level, situational and environmental factors such as funding, electronic information resources, university-industry relationships, the availability of books, professional journals positively influence the publication output of institutes. Additionally, factors such as lower teaching loads for researchers, research permissions, and attendance at international conferences enhance the publication productivity of the institute. (Wahid et al., 2022). A total of 160 different research institutions participated in the 1065 publications with institution information. It should be noted that an author may be affiliated with more than one institution, or a publication may be authored by individuals from different institutions. Of the total institutions, 86.87% (n = 139/160) were involved in fewer than 10 publications, and 13.12% (n = 21/160) were involved in 10 or more publications on the topic of machine learning in marketing. Table 5 provides information about the top 3 most productive institutions on the topic of machine learning in marketing, all of which are located in the United States of America. The institution with the highest number of publications on the topic (n= 15) is Carnegie Mellon University. The majority of the top 10 institutions were universities.

RANK	INSTITUTION	PUBLICATIONS
1	Carnegie Mellon University	15
2	University of Pennsylvania	14
2	Massachusetts Institute of Technology	14
3	University of Southern California	13
3	Indian Institute of Technology Delhi	13
3	Amity University	13
3	K L Deemed to be University	13

Table 5 Top 3 Most Productive Institutions in Machine Learning in Marketing
Source: Own elaboration

Keyword analysis

Keyword identification aids in discerning the theme or category of each publication. In the total count of 2358 publications on machine learning in marketing, 894 publications were identified with the keywords "Machine Learning" and/or "Marketing," showcasing a significant range of topics associated with machine learning in marketing publications. Table 6 presents the top 9 subject areas assigned by Scopus that are most frequently linked to the themes of machine learning in marketing publications. Figure 2 provides a legend of the names of the top 10 principal subject categories, accompanied by the total number of machine learning in marketing publications belonging to these categories.

The thematic category encompassing the most publications in the field of machine learning in marketing was "Machine Learning" with 1383 publications, followed by "Marketing" with 762 publications. Publications on "Machine Learning" are documented from the years 2001 to 2022, with the year 2021 being particularly prolific on this subject. As for "Marketing," publications began to emerge from the year 2003, with 2021 marking the year with the highest number of publications.

Figure 2 Keywords Cloud, created using WordCloud.com

RAKN	KEYWORD	PUBLICATIONS
1	Machine Learning	531
2	Marketing	363
3	Learning Systems	1
4	Commerce	47
5	Artificial Intelligence	105
6	Data Mining	38
7	Learning Algorithms	49
8	Sales	47
9	Forecasting	38

Table 6 Top 10 Keywords
Source: Own elaboration

Most utilized machine learning areas in marketing problems

Machine learning has a significant influence in the field of marketing, enabling deeper analysis and more precise decision-making. By employing machine learning algorithms, marketing professionals can scrutinize vast amounts of data to uncover patterns and trends, predict customer behaviors, and personalize advertising efforts. Upon reviewing the articles, it becomes evident that they are primarily focused on utilizing machine learning algorithms to address prediction and/or click-through rate enhancement in Facebook ads, sentiment analysis on Twitter, user behavior on Facebook, comments analysis on Facebook, popular topics on Twitter, music genre classification, and customer purchase prediction (see Figure 3). Likewise, advancements in AI, machine learning (ML), data science, and natural language processing (NLP) are enabling the development of conversational bots for a variety of applications, benefiting humans and contributing to the rapid growth in the number of chatbots. Currently, they play a positive role in shaping consumers' purchase intentions. (Liu & Duffy, 2023)
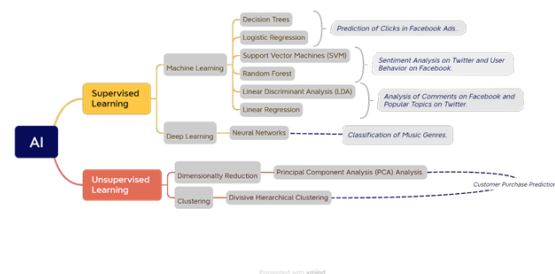

Figure 3 Most Utilized Algorithms in Digital Marketing, created using Xmind based on a self-generated table with 50 sample articles

Conclusions

In conclusion, Bibliometrics has become a fundamental tool for measuring the impact and evolution of scientific production, enabling the understanding of trends within specific subjects. In the case of machine learning in marketing, significant growth has been observed during the year 2021, driven by the impetus of e-commerce due to the pandemic. Additionally, it is evident that countries providing substantial support to scientific research continue to lead in terms of productivity in publications, such as the United States of America, India, China, and Canada.

In this regard, the present bibliometric analysis emerges as a valuable tool for analysis and decision-making within the scientific and technological sphere concerning machine learning in marketing. Furthermore, it is observed that there is a greater utilization of machine learning techniques in marketing, with approximately 75% being supervised learning algorithms, while unsupervised and other algorithms comprise around 25%. This suggests that there is an inclination towards automated detection using data containing some form of event identification or behavior to study. On the other hand, the detection of patterns without labels or identifiers remains a challenge, as the associated variable for the identified label is unknown. Such patterns indicate the presence of a phenomenon, yet they do not inherently explain it or provide actionable insights to the user.

References

Christian Federico Rueda-Clausen Gómez, Cristina Villa-Roel Gutíerrez, & Christian Eduardo Rueda-Clausen Pinzón,. (2005). Indicadores bibliométricos: Origen, aplicación, contradicción y nuevas propuestas. 8(1).

Eduardo Luis De Vito. (2006). Algunas consideraciones en torno al uso del Factor de Impacto y de la Bibliometría como herramienta de evaluación científica. https://www.redalyc.org/articulo.oa?id=382138365005

Larivière, V., & Sugimoto, C. R. (2019). The Journal Impact Factor: A Brief History, Critique, and Discussion of Adverse Effects. En W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), Springer Handbook of Science and Technology Indicators (pp. 3-24). Springer International Publishing. https://doi.org/10.1007/978-3-030-02511- 3_1

Mes, Año Vol.1 No.1 1-15-[Uso de ECORFAN]

Liu, L., & Duffy, V. G. (2023). Exploring the Future Development of Artificial Intelligence (AI) Applications in Chatbots: A Bibliometric Analysis. International Journal of Social Robotics, 15(5), 703-716. https://doi.org/10.1007/s12369-022-00956-0

Liu, Y., & Yang, S. (2022). Application of Decision Tree-Based Classification Algorithm on Content Marketing. En

JOURNAL OF MATHEMATICS (Vol. 2022). HINDAWI LTD. https://doi.org/10.1155/2022/6469054

Muhammad Rosyihuddin & Sudarmiatin. (2023). Analysis of New Terms in Digital Marketing Innovation Research Themes: Bibliometric Analysis. Indonesian Journal of Business Analytics, 3(2), 409-422. https://doi.org/10.55927/ijba.v3i2.3341

Tatiana Alexandra Escorcia Otalora. (2008). El análisis bibliométrico como herramienta para el seguimiento de publicaciones científicas, tesis y trabajos de grado. https://repository.javeriana.edu.co/handle/10554/8212

Wahid, N., Warraich, N. F., & Tahira, M. (2022). Factors influencing scholarly publication productivity: A systematic review. Information Discovery and Delivery, 50(1), 22-33. https://doi.org/10.1108/IDD-04-2020-0036.