Social Network Analysis and Mining (2022) 12:109
https://doi.org/10.1007/s13278-022-00916-6

ORIGINAL ARTICLE

Bibliometric analysis of the published literature on machine learning in economics and econometrics

Ebru Çağlayan Akay[1] · Naciye Tuba Yılmaz Soydan[1] · Burcu Kocarık Gacar[2]

Received: 5 December 2021 / Revised: 21 June 2022 / Accepted: 24 June 2022

Abstract
An extensive literature providing information on published materials in machine learning exists. However, machine learning is still a rather new concept in the fields of economics and econometrics. This study aims to identify different properties of published documents about machine learning in economics and econometrics and therefore to draw a detailed picture of recent publications from bibliometric analysis perspectives. For the aim of the study, the data are collected from the publications indexed by Web of Science and Scopus databases from the period 1991 to 2020. Inthe study, the data have been illustrated by VOSviewer for science mapping. The analysis of variance has also been used to identify the links between the number of citations of articles and years. The findings obtained provides information about the studies on machine learning in the relevant field conducted in the past, as well as providing an opportunity to gain knowledge about the researched area by shedding light on what the future research areas would be. There is no doubt that it attracts attention has increased significantly on machine learning in the field of economics and econometrics and academic publications on machine learning in the relevant field have increased over the last decade.

Keywords Bibliometric analysis · Economics · Econometrics · Machine learning · Science mapping · Scopus · Web of science

JEL Classification C45 · A10 · C10

1 Introduction

In recent years, the increase amount of data produced and shared has led to interest in big data and machine learning analysis. Big data and machine learning, which have frequently been used in the fields of biology, genetics, engineering, and astronomy, are still a rather new concept in the fields of economics, particularly econometrics. However, as in many other fields, our opportunities to collect big data with respect to variables having different measurements in the fields of economics and econometrics are gradually increasing. The features of big data, such as sample size and high dimensionality, lead to the requirements for and interest in methods and algorithms provided by these disciplines, such as machine learning for research purposes in the field of econometrics. These data which allowing greater focus on wider and more detailed analysis of economic activity and interaction will influence question types of economists and will provide more information in the future (Eivan and Levin 2014). Nowadays, many researchers are more interested in machine learning algorithms and data-driven approaches in applied economics and econometrics, even if the scope and purpose in machine learning are different. While econometrics has generally focused on explanation and more interested in causal inference, machine learning has focused more on prediction. Prediction and causal inference have been treated as two separate problems. To figure out the contents of the big data clearly and to understand the difference between machine learning and econometrics

✉ Burcu Kocarık Gacar
burcu.kocarikgacar@deu.edu.tr

Ebru Çağlayan Akay
ecaglayan@marmara.edu.tr

Naciye Tuba Yılmaz Soydan
tuba.yilmaz@marmara.edu.tr

1 Department of Econometrics, Marmara University, Istanbul, Turkey

2 Department of Econometrics, Dokuz Eylul University, Izmir, Turkey

are so important in order to use big data in econometrics. Econometrics have developed a body of insights on topic ranging from causal inference to efficiency that has not yet been incorporated into mainstream machine learning, while other part of machine learning has overlap with methods that have been used in social sciences for many decades. Many traditional methods do not perform well for big data. New statistical thinking and computational methods are needed to handle challenges of big data (Fan et al. 2014). While machine learning is particularly fallacious in the topics such as dimension reduction, model selection, data analysis, there are developments in the perpectives of applied economics and econometrics issues related to causal inference. Athey (2015) provides a brief overview of how machine learning relates to causal inference. In their studies, Athey and Imbens (2015) contribute to causal inference with machine learning in the field of econometrics. In the other study Athey (2018) defines machine learning and describes its strengths and weaknesses, and contrasting machine learning with traditional econometrics tools for causal inference. She also provides an overview of the questions considered and early themes of the emerging literature in econometrics and statistics combining machine learning and causal inference.

There are a few studies in the field of economics and econometrics, while there are many studies on big data and machine learning. It is not surprising that the literature is still limited because the development in big data econometrics are very new. Fortunately, there is some highly interesting literature available about big data and machine learning in econometrics, even if the field is relatively new and a lot of papes are still work-inprogress. These econometrics papers and talks during the last few years signal a sparked interest in machine learning. For example, Varian (2014) provides and introduction to many more novel tools and tricks from machine learning, such as decision trees and cross-validation. Einav and Levin (2014) describe big data and economics more broadly and summarize the challenges of big data in economics. There are also publications that openly advocate that the big data and machine learning will be an important power for the future of economics and econometrics (Athey 2017, 2018). Over the past few years, it has been observed the use of big data and machine learning in econometrics is important, new methods are developed, meetings and conferences are organized, academic studies and applications are made and these are increasing (Taylor et al. 2014; Varian 2014; Sengupta 2015).

The study aims to investigate the publication documents including journal articles, conference papers, reviews, books, book chapters, and editorial materials for machine learning in the fields of economics and econometrics literature and to provide evidence for the state, development, and impact of machine learning in the relevant fields employing the bibliometrics analysis. To examine trends in publications and co-authorship status, the authors collected data from the publications indexed on the Web of Science (WoS) and Scopus, which are the two largest databases, for the period 2010–2020. These two databases have made acquiring large volumes of bibliometric data relatively easy (Donthu et al. 2021). In the study, we examined these two databases separately and compared their similarities and differences to have more information on the publications in relevant fields.

To the best of our knowledge, the studies employing bibliometric analysis on machine learning in the fields of economics and econometrics have never been examined earlier. The bibliometric analysis will be new in these relevant fields and provide information about the strengths and weaknesses of the research area. This study contributes to the literature by providing current information on the state of Machine Learning in Economics and Econometrics research and draws a detailed picture of recent publications from bibliometric analysis perspectives. This information provided will shed light on the improvement of machine learning in the fields of economics and econometrics. This study also visualizes scholarly networks, and developments to identify future exploring research collaboration among researchers and research areas.

The research process of the study is as follows. (i) after preprocessing publications were exported from WoS and Scopus. 3692 publications in total were analyzed based on annual publications, and the publication of the most productive countries, institutions or journals and authors. The most effective studies and researchers in the fields of interest were determined by descriptive bibliometric analysis. (ii) by the use of VOSviewer, which is the most frequently used professional science mapping/social network tool, the co-citation, co-authorships, co-occurrence, bibliographic coupling networks of countries, institutions or journals, authors, and papers are depicted. The cooperation networks of countries, institutions or journals, and authors are illustrated by VOSviewer and the strongest collaborative relationships were listed. (iii) the analysis of variance (ANOVA) was performed to investigate the links between the number of citations of articles and years.

The rest of the paper is organized as follows. After a brief introduction, we present the literature on bibliometric analysis and science mapping in the machine-learning literature. Sections 3 and 4 outline the methodology and research questions and data, respectively. The findings section presents the results of the bibliometric and science mapping. Finally, the last section concludes remarks.

2 Literature

This bibliometric analysis is the most extensive analysis in which mathematical and statistical methods are used to effectively measure, analyze and assess the bibliographic information of studies published on a specific subject such as Goyal and Kumar (2021), Xu et al. (2018), Linnenluecke et al. (2017) and Durisin and Puzone (2009), Bonilla et al. (2015), Castillo-Vergara et al. (2018), Wang et al. (2020a, b), Merediz-Sola and Bariviera (2019) in finance and economics, Elie et al. (2021), Rosokhata et al. (2021), Bortoluzzi et al. (2021), Sarkodie and Owusu (2020), in renewable energy, Ellegaard and Wallin (2015), Merigo and Yang (2017), Fahimnia et al. (2015), Zupic and Cater (2015) in management, Farrukh et al. (2020), Ferreira et al. (2011), Kumar et al. (2021a, b) in business strategy, Backhaus et al. (2011), Miskiewicz (2020), Donthu et al. (2020), Donthu et al. (2021), Gao et al. (2021) and Hu et al. (2019) in marketing, Julius et al. (2021), Sönmez (2020) in education. Today, despite the existence of numerous studies in different fields that use big data and machine learning algorithms it is observed that there are a few studies that examine big data and machine learning through bibliometric analysis. Among these studies, for instance, while Belmonte et al. (2020) examined big data and machine learning simultaneously in their bibliometric analysis, Alonso et al. (2018), Dhamija and Bag (2020), Tran et al. (2019) focused on artificial intelligence, and Li et al. (2020), Ali et al. (2022), Bidwe et al. (2022), Mao et al. (2018), Nakhodchi and Dehghantanha (2020) focused on deep learning. Many researchers, including Mishra et al. (2018); Xian and Madhavan (2014); Liao et al. (2018); Ardito et al. (2019); Liu et al. (2019) and Rialti et al. (2019), Kalantari et al. (2017) have reviewed studies concerning big data. Some studies have also been conducted to examine the place and improvement of machine learning in different fields. For instance, Makawana and Jhaveri (2017); Stout et al. (2018); Bhattacharya (2019); Dos Santos et al. (2019); Salod and Singh (2020) and Wang et al. (2020a, b) revealed the place of machine learning in the field of health through bibliometric analysis. Linden et al. (2017) used both bibliometric and social networks analysis metrics. Perez-Aranda and Pelaez-Verdet (2021) have an application for social network mining. ANOVA was applied in some studies in different fields to examine the bibliometric results. For instance, Kalantari et al. (2017); Pesta et al. (2018); Cortés-Sánchez (2020); Perez-Aranda and Pelaez-Verdet (2021); Demirkol et al. (2022) have applied ANOVA for various metrics.

As seen in the literature overview, the studies based on the bibliometric analyses of the machine-learning literature have been employed in numerous different fields. However, there are absent any studies on machine learning in the fields of economics and econometrics that have examined these analyses. Thus, we believe that this study will help identify issues and the field's development and will be a valuable tool for applied economists and econometrics who are interested in machine learning and will provide an essential contribution to literature.

3 Research questions

Following the aim of the study, the development of studies using machine learning in the fields of economics and econometrics were examined by descriptive and evaluator bibliometric analysis and were illustrated with science mapping. The study examined the publication numbers, document type, keywords, authors, and institutions that contributed to most machine learning research in the fields of economics and econometrics. In the study, we focused on the following questions designed.

- (i) What was the number of publication and how chanced the distrubution of these publications by document type about machine learning in the fields of economics and econometrics during research period?
- (ii) What is the frequency of keywords repeated in the studies of the relevant fields?
- (iii) What are the top contributors and publications in the relevant fields?
- (iv) Which institutions and authors have published the most articles on these fields?
- (v) How is the visualized intensity structure of the authors' network in the studies?
- (vi) Is there a significant relationship between the number of citations of articles and years?

The research questions to be answered in the study were performed using bibliometric analysis, citation, content analysis, science mapping, visualization with Voswiever, and ANOVA analysis. The descriptive information was provided for the publications and classified with headings such as the country, time, institution, and subject of the research.

4 Methods

4.1 Bibliometric analysis

The word bibliometrics, which was first used by Pritchard (1967), expresses a measurement unit of a book and/or document. Bibliometric analysis may generally be defined as a means used to identify patterns in big databases and to deduce information to explain undefined behaviors (Daim et al. 2005). This analysis is being used to determine the most effective studies and researchers in a specific field. *Descriptive bibliometric* results are able to be obtained in the case of making classifications, such as the country, time, institution, field, and subject of the research by determining the number of journals, articles, and papers published in the addressed subject. Moreover, it is also possible to obtain *evaluator bibliometric* results revealing the relationships established in the literature with respect to the examination of subjects, such as cited authors, considering their references (Nicholas and Ritchie 1978). The methodology used in the bibliometric analysis, as well as mapping and presenting the data visually along with figures, assist researchers in understanding the relationships between explanatory values and various aspects of the identified studies.

4.2 Science mapping

Science mapping which ensures the understanding of the complex relationship network among the analyzed units, analyzing the social interactions within a complex structure, and revealing the network structure across units by mapping them is frequently used in bibliometric studies. Science mapping connects the actors such as people, groups, institutions, or countries and the knots of the relationships between these actor pairs. The network structures in the subject, position of units within the network, and clustering may be identified through references and ascriptions or as a result of digitization/vectorization within the frame of specific measurements, for the purpose of understanding the involved complexity (Gürsakal 2009; Knoke and Yang 2020). Science mapping aims to build bibliometric maps that describe how scientific domains, research fields, or specific disciplines are conceptually, intellectually, and socially structured (Cobo et al. 2011). The findings of the science mapping could be used in the future for purpose of more powerful information generation (Yoopetch and Nimsai 2019). Science mapping has been used in studies of science citations, social mobility, class structure, and many other areas (Scott 1988; Newman 2001; Perianes-Rodríguez et al. 2010; Perez-Aranda and Pelaez-Verdet 2021).

Science mapping has different citation-based approaches such as citation, co-citation analysis, and bibliographic coupling. Citation analysis operates on the assumption that citations reflect linkages between publications that are formed when one publication cites the other (Appio et al. 2014; Pieters and Baumgartner 2002; Stremersch et al. 2007). Co-citation analysis assumes publications that are cited together frequently are similar thematically (Hjørland 2013; Donthu et al. 2021). Bibliographic coupling concentrates on the division of publications into thematic clusters based on shared references. Besides these three approaches, co-keyword and co-authorship analysis are also commonly used in science mapping. The first three approaches focus on publications, but co-keyword analysis examines the actual content of the publication itself and focuses on "author keywords" (Baker et al. 2020). Co-authorship analysis is widely used to understand and assess scientific collaboration patterns. It examines the interactions among scholars in a research field (Fonseca et al. 2016).

4.3 Visualization with VOSviewer

VOSviewer is a software tool for constructing and visualizing bibliometric networks developed by Van Eck & Waltman. These bibliometric networks may for instance include journals, researchers, or individual publications. They can be constructed based on citation, bibliographic coupling, co-citation, or co-authorship links. VOSviewer can create node-link maps from the bibliographic data of the documents and visualize the citation links (Waltman et al. 2010). Each link has a strength denoted by a positive numerical value and a high value means a strong link (van Eck and Waltman 2017, 2019; Su et al. 2021).

4.4 ANOVA

ANOVA is one of the most frequently used statistical methods developed by R.A. Fisher in the first half of the 1920s. It is used to analyze group means (by separating variation into components such as between groups and within groups) and to compare the means of two or more groups. The ANOVA compares the means between the two or more independent groups and determines whether any of those means are statistically significantly different from each other (Sthle and Wold 1989; Perez-Aranda and Pelaez-Verdet 2021). It uses the statistic F, which is the ratio of between and within group variances.

5 Data collection and data analysis

In the study, WoS and Scopus, which are the two most extensive databases, were selected as bibliometric data sources. WoS is a platform based on web technology created in 1960 and owned by "The Thomson Reuters Institute of Scientific Information". It has collected a wide range of bibliographic information, citations, and references from scientific publications in any discipline of knowledge-scientific, technological, humanistic, and sociological- since 1945. Scopus, officially named SciVerse Scopus, was introduced by Elsevier in November 2004 and is a bibliographic database of the scientific, multidisciplinary and international literature and has analyzed citations since 1996, providing a complete view of worldwide research production. In recent years, Scopus has taken steps to decrease its differences with WoS. The high competition between these databases motivates researchers to compare them. In the literature, when the studies indexed in WoS and Scopus are assessed, it is observed that both databases have superiority over the other in different subjects in terms of the number of publications and ascriptions (Aghaei Chadegani et al. 2013).

The data used in the study were compiled by a review of the documents indexed in these two databases (WoS and

Scopus) selected for economics and econometrics, machine learning, and big data between 2010 and 2020. The explanation with a flowchart about how collected data in our study and screening and constraints of the study are given in Fig. 1.

6 Results and discussion

The results of the analyses made in the study cover three parts: (i) the results of bibliometric analysis obtained by BibExcel used for indicating the specifications of the searched publications and (ii) science mapping ensuring the determination of different relationships through visualization. We used VOSviewer for this analysis. (iii) the results of one-way ANOVA to identify the link between the number of citations of publications and years.

6.1 Publications overview

From the period 2010 to 2020, the studies regarding machine learning in economics and econometrics, 3692 publications in total, 840 from WoS, and 2852 from Scopus, were analyzed. The number of studies published on machine learning in the fields of economics and econometrics obtained from WoS and Scopus for the period is shown in Fig. 2.

According to Fig. 2, it is clear that machine learning has not seemed so popular among economists and econometricians until 2014. From 2010 to 2015, the total number of studies in the fields of machine learning and economics, and econometrics are 120 and 441 for the WoS and Scopus databases, respectively. In 2014, there are publications that openly advocate that big data will be an important power for the future of economics and econometrics (for example, Varian 2014; Einav and Levin 2014; Fan et al. 2014 among others). It is observed that increased interest in this field was actualized following these publications and that increased interest occurred in terms of the number of studies made since 2015 in both databases. The figure shows that from 2016 to 2020, the total number of studies are 720 and 2411 for the WoS and Scopus databases, respectively. In these periods, studies of econometrics in this area are mainly based on the involvement of causal relations in machine learning algorithms (Athey 2015, 2017 and 2018; Wager and Athey 2017; Athey and Imbens 2017 among others). All these findings indicated that many researchers use machine learning algorithms and approaches to econometrics, despite the different scope and purposes of machine learning.

The distribution of data by publication types including journal articles, conference papers, reviews, books, book chapters, and editorial materials for machine learning in the

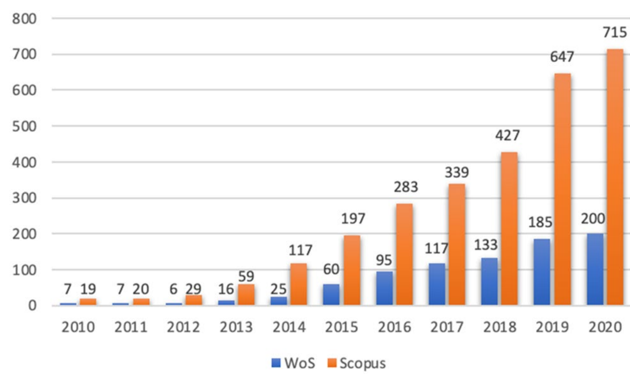


Fig. 2 Number of publications by year

Fig. 1 Flowchart according to the documents identification

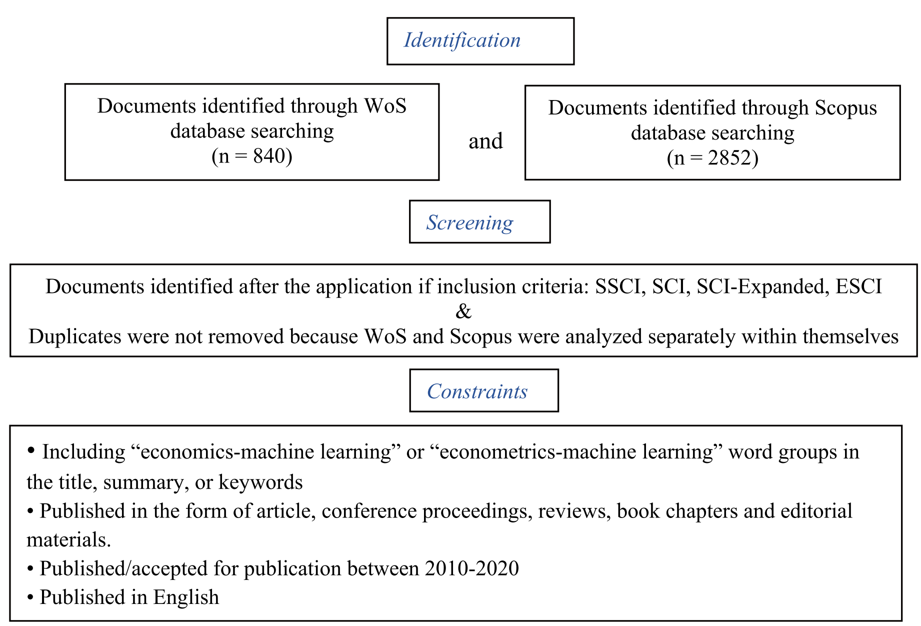


Table 1 Number and distribution of publications by document type

Document type	Web of science		Scopus	
	2010–2020		2010–2020	
	Number	Percent (%)	Number	Percent (%)
Journal articles	532	63.3	1263	44.3
Conference papers	223	26.6	1245	43.7
Reviews	59	7.0	161	5.6
Others*	26	3.1	183	6.4
Total	840	100	2852	100

*It expresses books, book chapters and editorial materials

fields of economics and econometrics literature is presented in Table 1. According to the table, it is observed that in the period 2010–2020 the number of articles ranked first for both databases, and the number of conference papers was ranked second for both databases.

6.2 Countries overview

The results of the identification of the most frequently published articles in terms of the countries of the institutional affiliation of the authors are reported in Table 2. Fifteen countries published the highest number of articles and the number of ascriptions made to them is presented in Table 2.

Figure 3 shows that majority of institutions or universities in the USA have the highest citation whereas this is followed by the China and UK in both databases. According to Fig. 3a and b, two main clusters (colors) occurred among the countries. The color red in the network intensity map indicates the country with the highest number of ascriptions which is the US. Thus, the first cluster in the subject consists of the US and other countries related to it. Moreover, it is observed that China and the UK constitute the second cluster (color), shown in a lighter color when examined in terms of the intensity of ascriptions.

Table 2 the countries that published the most articles and number of ascriptions made to them

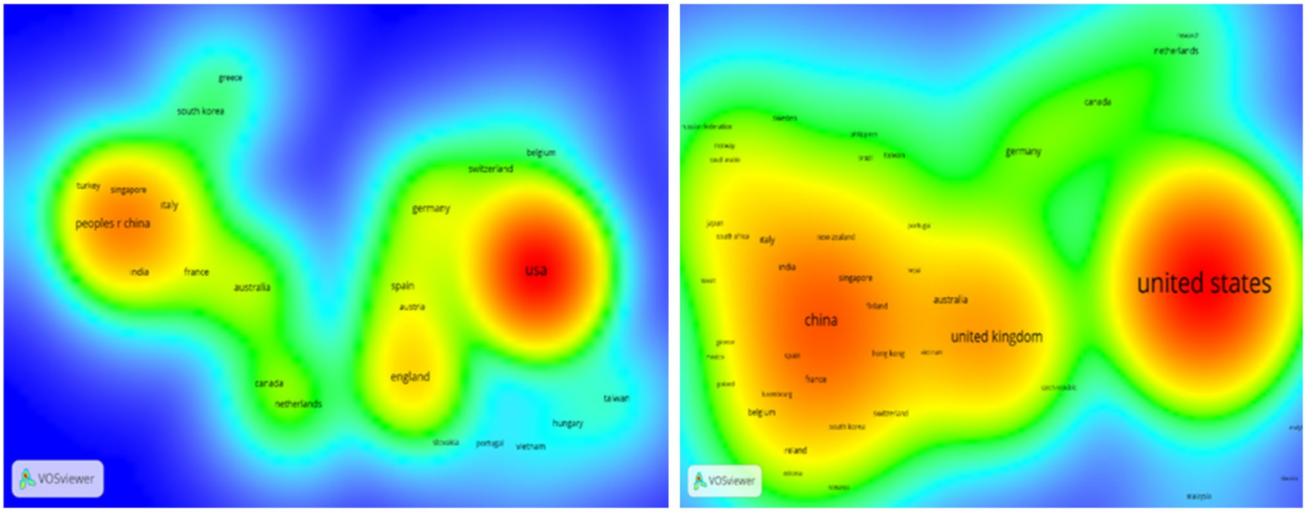
Web of science			Scopus		
Country	Number of documents	Number of citations	country	Number of documents	Number of citations
US	311	4260	US	799	12,222
China	119	1212	China	635	4580
UK	85	1049	UK	223	3861
Germany	51	467	India	148	648
Australia	42	599	Germany	131	1315
Italy	38	268	Canada	91	1056
Spain	34	175	Italy	89	1118
France	27	391	Russia	86	294
Canada	26	177	Australia	84	1265
India	24	41	S. Korea	71	602
Russia	23	22	Spain	69	567
S. Korea	20	158	France	61	790
Netherlands	19	63	Japan	59	585
Switzerland	19	85	Switzerland	46	616
Japan	17	48	Brazil	43	265
Taiwan	16	319	Netherlands	43	1028
Singapore	14	281	Taiwan	42	633
Austria	12	105	Hong Kong	41	612
Sweden	11	140	Greece	37	245
Turkey	10	37	Singapore	36	696

6.3 Institutions and organizations overview

The institutions and organizations that published the most articles in 2010–2020 and the number of ascriptions made to them were examined, and the top twenty institutions are tabulated according to the number of documents in Table 3.

From 2010 to 2020, the institution that published the most articles and that had the highest number of ascriptions on WoS (840) and Scopus (2852) was the University of California (United States (US)). When evaluated in terms of the numerous documents indexed in WoS, Harvard University, Oxford University, and MIT followed the University of California (US). Moreover, when evaluated in terms of documents indexed in Scopus, it was observed that this order was different. According to Scopus, Stanford University, the Chinese Academy of Sciences, and Harvard University followed the University of California. The principal institutions among all considered institutions are provided in Table 3.

Those institutions ascribed to the most in WoS were the University of California (US), and MIT (Massachusetts Institute of Technology). Moreover, it is observed that Stanford University (US), and Google Inc. (US) have a high rate of ascriptions with a smaller number of publications compared to other institutions. Those institutions ascribed to the most in Scopus were the University of California (US), Stanford University (US), Harvard University (US), and the Chinese Academy of Sciences (China). Moreover, it should be noted that Princeton University (US) and Carnegie Mellon University (US) have high rates of ascriptions with fewer publications compared to other institutions. In this context, it is meaningful to mention that factors such as scientific productivity and scientific publication performance are highly related to the number of ascriptions.



a *Web of Science* **b** *Scopus*

Fig. 3 Network structure relationship among countries

Table 3 Institutions and organizations that published the most articles in 2010–2020 and the number of ascriptions made to them

Web of science				Scopus			
Organization	Doc	Cit	Country	Organization	Doc	Cit	Country
Univ. of California	21	590	US	Univ. of California	85	3843	US
Harvard Univ	15	340	US	Stanford Univ	62	2536	US
Oxford Univ	13	243	UK	Chinese Academy of S	56	582	China
MIT	12	510	US	Harvard Univ	46	1691	US
Stanford Univ	12	378	US	Oxford Univ	39	536	UK
Carnegie Mellon Univ	12	131	US	Univ. of Pennsylvania	27	210	US
Univ. of Pennsylvania	11	121	US	Carnegie Mellon Univ	22	438	US
Univ. Illinois	11	34	US	Beijing Jiaotong Univ	21	109	China
New York Univ	9	333	US	Univ. of Texas	20	90	US
Univ. of Chicago	9	146	US	Imperial College London	19	317	UK
Univ. College London	8	287	UK	Univ. of Washington	19	253	US
Univ of California San Diego	7	210	US	Univ. Illinois	18	137	US
Chinese Academy of S	7	155	China	Princeton Univ	17	581	US
Rutgers Univ	7	30	US	Rutgers Univ	16	80	US
Yale Univ	6	184	US	Univ. of Chicago	15	279	US
Univ. Southampton	5	290	UK	National Research Univ	15	15	Russia
Google Inc	4	355	US	MIT	11	116	US
Imperial Collage London	4	190	UK	Univ. of Cambridge	10	238	US
National Bureau of Economic Res	3	190	US	Duy Tan Univ	10	34	Vietnam
National Taiwan Univ	2	193	Taiwan	Yale Univ	10	300	US

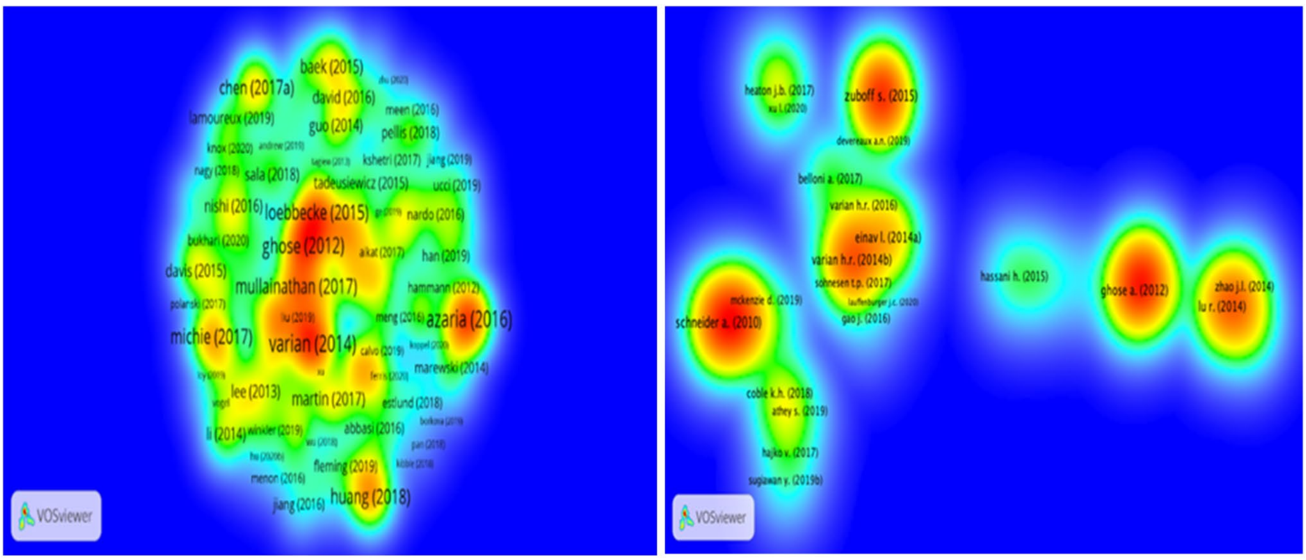
Doc. and Cit. represent "number of documents" and "number of citations", respectively. The order in the table is made from the highest number of documents to the least

6.4 Author overview

In Fig. 4, the visualized intensity structure of the network in studies with the highest interaction rate is observed according to the name of the first author. The intensity structure indicating the relationship of the most ascribed authors with other authors is observed in Fig. 4. The color red in the network intensity map indicates the document with the highest number of ascriptions, while the color green indicates the document with the less number of ascriptions. Some documents/authors do not appear on the network due to overlaps.

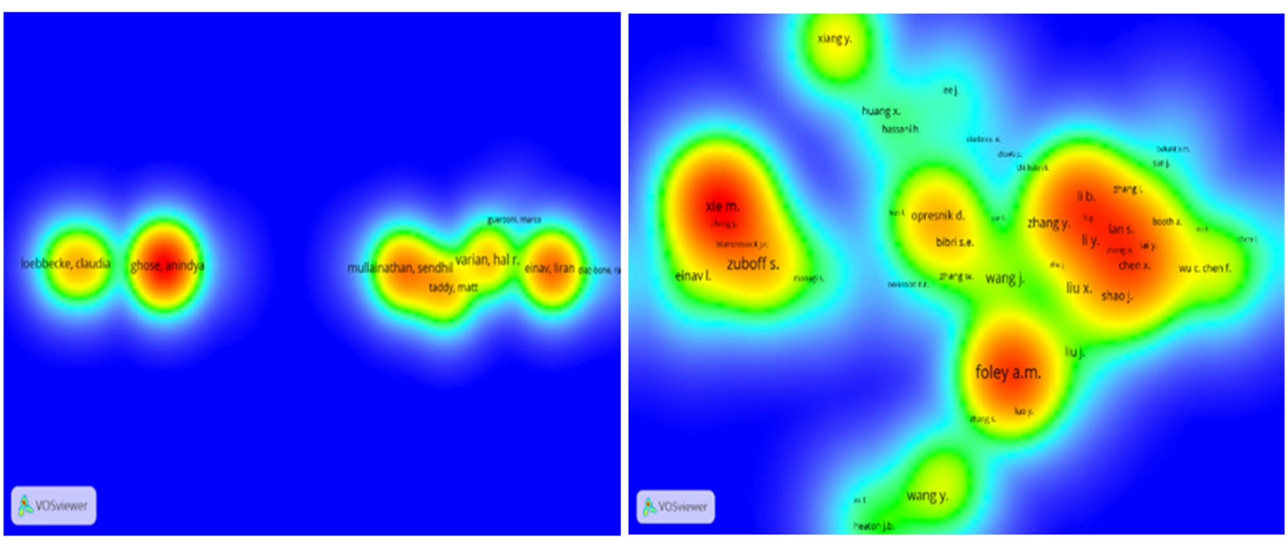
According to Fig. 4, while the studies of Varian (2014), Mullainathan (2017), Ghose (2012), Loebbecke (2015), Azaria (2016), and Huang (2018) come to the forefront in the WoS database, those of Zuboff (2015), Varian (2014), Schneider (2010), Ghose (2012), Zhao (2014) and Lu (2014) come to the forefront in the Scopus database.

Figure 5 illustrates that while Varian, Mullainathan, and Taddy form a cluster due to their mutual ascription relationships in the WoS database, Ghose forms a cluster as a reference to other authors. According to the Scopus database, it is observed that the network of ascriptions is being formed by more authors. It can be said that in the Scopus database, the co-citation relations network includes more than the WoS database. These authors are Xie and Zuboff for the left cluster (authors in red color and larger font), Zhang, Liu, Shao, and Chen mainly for the right cluster (authors in red and larger fonts), and Foley for the following cluster (authors in red and larger fonts).



a *Web of Science* **b** *Scopus*

Fig. 4 Intensity map among documents (First Author's Name Abbreviated)



a *Web of Science* **b** *Scopus*

Fig. 5 Intensity map among authors

The authors who are ascribed to the most and with the highest number of publications in both databases are shown in Table 4. According to this, while Varian (2014) is coming to the forefront as the author who is ascribed to the most the WoS database, Foley, Leahy, Marvuglio, and Mckeogh (2012) are the authors ascribed to the most in the Scopus database. Moreover, it is observed that the author with the highest number of publications is Y. Wang, with 29 publications.

7 Keywords overview

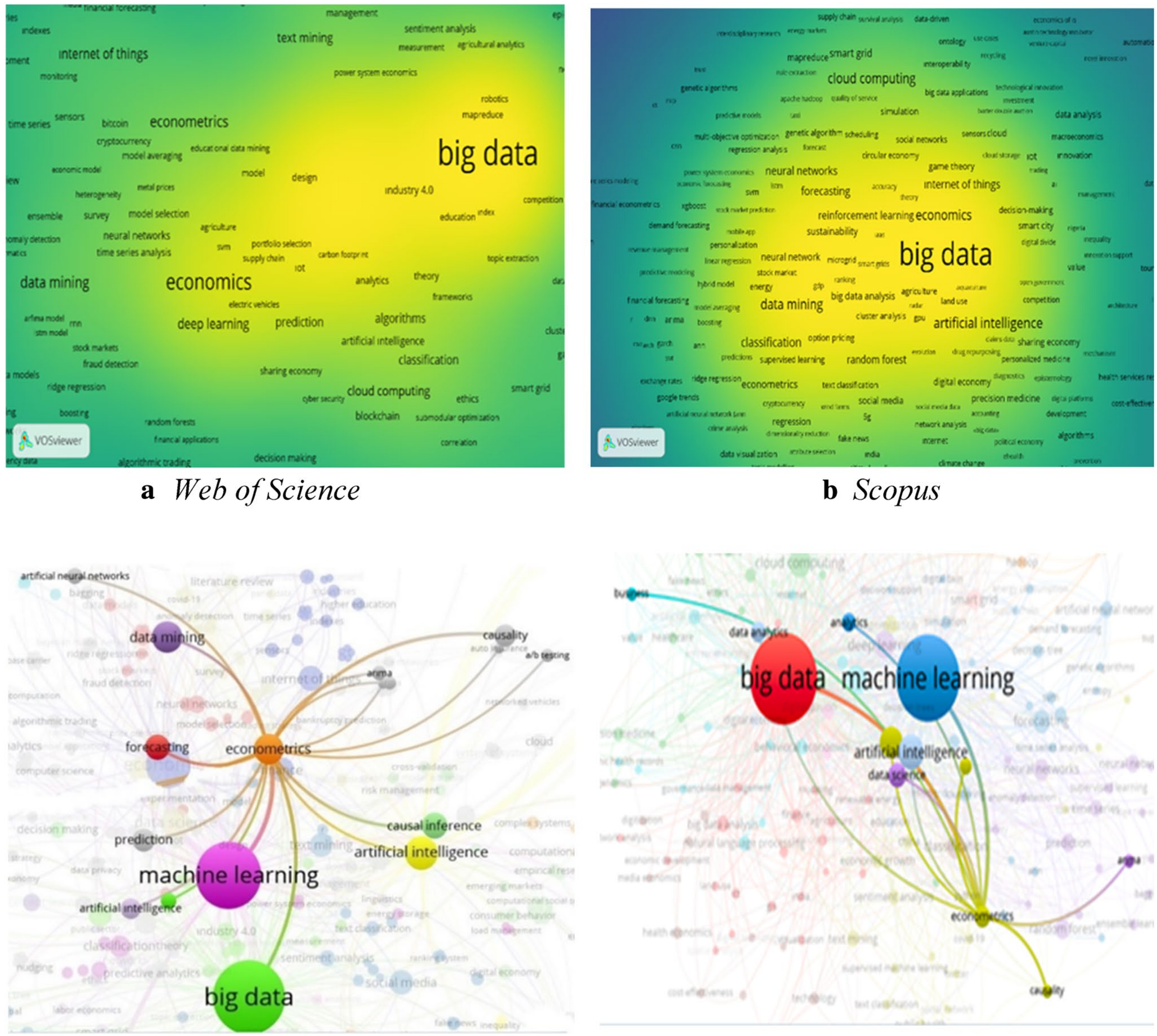
A keyword analysis was performed in the titles and abstracts of the papers to explore the research topics. Figure 6 presents the density of the words used in the titles and the change over time. When the use of keywords is evaluated (provided in "Appendix 1" in detail) for each year, it is observed that in the studies indexed in both WoS and Scopus, the keywords “economics and econometrics”, “machine learning”, “big data”, and “data mining/data analysis/data science/data analytics” are used together and that the interest in these words is increasing each year. Thus, it can be said that the research concerning machine learning in the fields of economics and econometrics is rapidly increasing. The abundance of the word “big data” especially used in the studies in this field is drawing attention ("Appendix 1"). This status is observed in the network intensity map among the words in Fig. 6a and b. From 2010 to 2015, the start of the joint use of the words “economics and econometrics”, “machine learning”, and “causality-causal inference” and the increase in the frequency of such usage in the following years indicate the presence of actual research on these subjects. From 2016 to 2020, it is found that the words “artificial intelligence”, “cloud computing”, “deep learning”, “neural networks”, “internet of things”, and “reinforcement learning”, are provided by the related technologies in addition to all these, “text mining”, “support vector machine”, “algorithms”, “time series analysis”, and “optimization” are frequently associated.

The different colors of the nodes are to prevent overlapping large and small nodes from mixing with each other. The sizes of the nodes, not the colors, represent their importance. In the keywords section of the studies, these words were found to be related because they were often combined together. Because they have large nodes, they are brought to the fore and the edges are shaded. In Fig. 6, the nodes in the network include keywords, while the links establish relationships or flow

Table 4 Ranking of authors with the highest number of publications and who are ascribed to the most

Web of science						Scopus					
Number of ascriptions			Number of documents			Number of ascriptions			Number of documents		
Author	Docs	Cit	Author	Docs	Cit	Author	Docs	Cit	Author	Docs	Cit
Varian, H. R	2	353	Peysakhovich, A	4	13	Foley, A. M	1	709	Wang, Y	29	324
Azaria, A	1	350	Mullainathan, S	3	186	Leahy, P. G	1	709	Zhang, Y	27	317
Ghose, A	1	226	Athey, S	3	171	Marvuglia, A	1	709	Wang, J	22	351
Li, B	1	226	Hoi, Steven C. H	3	124	Mckeogh, E. J	1	709	Li, Y	21	378
Ipeirotis, P. G	1	226	Li, B	3	124	Azaria, A	1	611	Chen, Y	18	108
Mullainathan, S	3	186	Swanson, N. R	3	19	Ekblaw, A	1	611	Wang, L	18	144
Spiess, J	1	183	Varian, Hal R	2	353	Lippman, A	1	611	Liu, Y	17	110
Rust, R. T	2	174	Rust, Roland T	2	174	Vieira, T	1	611	Liu, J	16	242
Loebbecke, C	1	173	Imbens, Guido W	2	153	Zuboff, S	1	552	Li, X	15	134
Picot, A	1	173	Akter, S	2	146	Babuška, R	1	537	Liu, X	15	326
Athey, S	3	171	Shi, Y	2	76	Buşoniu, L	1	537	Wang, S	15	154
West, R	1	171	Acquisti, A	2	73	De Schutter, B	1	537	Wu, J	14	66
Yardley, L	1	171	Hu, J	2	24	Ernst, D	1	537	Zhang, J	14	55
Huang, M.-H	1	167	Zhang, Y	2	24	Bates, D. W	3	459	Lee, J	13	94
Huang, G.Q	1	161	Vieira, T	1	353	Escobar, G	1	444	Li, J	12	30
Lan, S	1	161	Ghose, A	1	226	Ohno-Machado, L	1	444	Li, Z	12	16
Newman, S. T	1	161	Li, B	1	226	Saria, S	1	444	Wang, H	12	31
Zhong, R. Y	1	161	Ipeirotis, P. G	1	226	Shah, A	1	444	Yang, Y	12	182
Imbens, G. W	2	153	Spiess, J	1	183	Varian, H. R	3	436	Zhang, L	12	74
Einav, L	1	147	Loebbecke, C	1	173	Xie, M	4	395	Liu, H	11	76

Doc. and Cit. represent “number of documents” and “number of citations”, respectively. The number of ascriptions column in the table are listed from the highest number of citations to the least and the number of documents column are listed from the highest number of documents to the least



a *Web of Science* **b** *Scopus*

c *Web of Science* **d** *Scopus*

Fig. 6 The most frequently used keywords (2010–2020)

between the nodes. Once the collaboration between keywords increases, the link becomes thicker. In this study, since the studies on machine learning in economics and econometrics were examined, attention was paid to the association of all the keywords obtained with these keywords. Therefore, the frequently used words among these words were tried to be filtered. Keywords used directly together with these keywords are given in Fig. 6c and d by zooming. Where the sizes of the nodes representing the keywords symbolize the frequency of use of the keywords. The network structure among the keywords from the publications on machine learning in the field of economics and econometrics is observed in Fig. 6c and d. The use of the words "health economics", "finance", "industry 4.0", "behavioral economics", "business intelligence", "economic growth", "digitalization", "digital economy", "sharing economy", and "e-commerce", "blockchain", "sustainability", and "sentiment analysis", which may be deemed within the scope of economics, is also frequently observed. The words "classification", "clustering", "forecasting", and "prediction", which are within the scope of machine learning techniques, are among the other significant words used in this field. Moreover, it is observed that the word "social media" is also included in these studies. In addition, it was determined that along with the gradually increasing research interest, the words "financial econometrics", "spatial econometrics", "machine learning econometrics", "Bayesian econometrics", and “structural econometrics” began to be used, along with machine learning, within the scope of econometrics.

The frequency of the keywords analyzed in the different periods clearly shows that there is an increased interest in machine learning methods in the fields of economics and econometrics. The findings of the analysis show that the keywords “causal inference” and “causality” have been mentioned in machine learning publications after the year 2014.

7.1 The results of ANOVA

ANOVA was performed to answer the question "Is there a significant increase in the average number of citations of Machine Learning in Economics and Econometrics studies over the years?". Therefore the years between 2010 and 2020 formed groups (for Wos and Scopus). The statistical significance of the difference between the citation numbers of the keywords by years was tested. The changes in econometrics, economics, and machine learning keywords in the databases over the years were compared. Three ANOVAs were conducted in the study. Table 5 presents the output of the ANOVA which show that for keywords of documents from 2010 to 2020 ("Appendix 1") and according to citation on WoS and Scopus databases together. ANOVA F test was applied to check the significance. The null hypothesis states that no real difference exists between the tested groups. Differences were considered significant if probability value (p value) is less than the 0.05 alpha level.

The ANOVA results show that there was a significant difference in the average number of citations of machine learning in economics ($F = 5.269$, $p < 0.05$) and machine learning in econometrics by years ($F = 4.110$, $p < 0.05$). Moreover, there was insignificant difference in the average number of citations of machine learning studies by years ($F = 2469$, $p > 0.05$). Findings of the ANOVA indicated that there is a significant increase in the number of citations has increased the number of studies in this field and attracted attention in the fields of economics and econometrics.

8 Implications, limitations, and further consideration

We explore how our findings answer the research questions and summarise the implications for practice. We also highlight some of the limitations of our study, and then we offer recommendations for further research.

8.1 Theoretical and practical implications

Big data is extensively being used in the fields of astronomy, biology, social sciences, and genetics, and, thus, machine learning is beginning to play a determinant role in many different fields through the progress of studies on the health sector, marketing, artificial intelligence, forecasting, with the focus being the financial and banking sectors. Since the reason such as the big data and machine learning in the field of economics and econometrics is relatively new and the machine learning algorithms are rarely included in causal inference, economists and econometricians have avoided the concepts of machine learning and big data. However, it can be observed that economics journals have begun to carve out a place for machine learning studies in recent years.

Using the bibliometric analysis, the data obtained from the WoS and Scopus databases were examined in the period 2010–2020 covering the subjects of "economics and machine learning", "econometrics and machine learning", and "economics and big data", and "econometrics and big data". The findings of the descriptive bibliometric analysis brought out the important information in the relevant fields such as the changes in the studies based on the year published, the author publishing the highest number of articles on the relevant subject, the distribution of publications across countries, and institutions, most repeated keywords, and improvements arising in time. Moreover, through science mapping and relationships among authors, the forming of network structures with each other (interacting), countries, and keywords were mapped and visualized. It was

Table 5 Results of ANOVA

ANOVA	Average degree	Sum of Squares	df	Mean Square	F	p value
Economics	Between Groups	3834.455	10	383.445	5.269	.006
	Inside Groups	800.500	11	72.773		
	Total	4634.955	21			
Econometrics	Between Groups	237.273	10	23.727	4.110	.014
	Inside Groups	63.500	11	5.773		
	Total	300.773	21			
Machine Learning	Between Groups	24,788.364	10	2478.836	2.469	.077
	Inside Groups	11,044.000	11	1004.000		
	Total	35,832.364	21			

ANOVA is used to by separate variation into components such as between groups and within groups. The total variation is the sum of the between groups variation and within groups variation for each group. *df* degrees of freedom found that the relationship between causality and big data and machine learning in the relevant fields has een established along with the increase in the number of studies in recent years, especially since 2018. It was concluded that in the field of econometrics, machine learning studies are related to artificial intelligence and forecasting studies.

Our findings show that the largest number of studies each year was performed in the US. In addition, when these studies were evaluated, it was observed that the tendency of subjects of machine learning, currently attracting attention in the fields of economics and conometrics, has rapidly increased and that these studies have been associated with the keywords such as “financial econometrics”, “behavioral economics”, and “experimental economics” as well as “causality” and “causal effect”. It also observed that keywords related to finance and economics (for example, industry 4.0, business intelligence, economic growth, digital economy, sharing economy, e-commerce, blockchain, and sustainability, among others) are frequently used along with machine learning. Consequently, the findings obtained in this study shed light on the improvement of machine learning in the fields of economics and econometrics.

8.2 Limitations of the study

In the study, before bibliometric analysis, machine learning studies on both economics and econometrics were filtered out from WoS and Scopus databases. Unfortunately, search results yielded very limited publications and these publications were too few to be examined in comparison. This can be due to the newness of the relevant field and the lack of some keywords and combinations of keywords, useful to integrate the database. One of the constraints that we faced during the study was the inability to search only by a “keyword” on the WoS database compared to Scopus. For this reason, the analyses were actualized as covering the title, abstract, and keywords of the study by selecting “topic” for both databases. Thus, the search results were filtered to include articles on both "machine learning in economics," and "machine learning in econometrics", separately, and a total of 3692 publications are examined from bibliometric analysis perspectives. We have also faced other issues directly linked to the newness of the field. Some articles on relevant fields even if published in journals are not included in our study due to the process of databases. We believe that it may be more useful in future studies whether the number of studies on these relevant fields increases enough.

8.3 Further considerations

We focused on only the most frequent keywords, and relations between them in the study. For future research directions, some new specific scientific keywords such as the hybrid model, which combines econometric methods and machine learning algorithms, could be included by the authors in their papers to obtain more information on the relevant field. The findings of our study clearly show that the tendency of subjects of machine learning, currently attracting attention in the fields of economics and econometrics, and the publication based on these fields has rapidly increased. Especially, our findings indicated that the publications have been associated with the keywords such as “financial econometrics", "Bayesian econometrics", "structural econometrics", “behavioral economics”, and “experimental economics”, recently. To provide more extensive information on the relevant fields, we suggest that future research could focus on sub-branches of the relevant fields such as specific keywords indicated above.

9 Conclusion

Big data has the potential to permit a better measurement of economic outcomes and allows to focus on a more detailed analysis of economic activity and provide more information in the future. This kind of data is often analyzed by using machine learning algorithms due to the features of big data. Employing machine learning methods can provide methodological and practical advantages over classical statistical methods. It is clear many researchers are more interested in machine learning algorithms approaches in applied economics and econometrics recently.

The present paper is the first study to provide extensive information to researchers regarding the status and future of academic publications in machine learning in the fields of economics and econometrics using bibliometric analysis and scientific mapping. We also performed ANOVA analysis of publications.

The study draws a detailed picture of recent publications in machine learning in the relevant fields. The findings of the study indicate that machine learning methods are still not employed by many researchers in the field of economics and econometrics. But the academic publication in the last few years clearly demonstrates the growing interest in big data and machine learning in econometrics. This growing interest has led to the use of big data and machine learning in the relevant fields.

We believe that this study could provide benefit for the future of big data and machine learning studies in econometrics, becoming more accepted by more researchers, becoming more widely known, and developing this particular area of research and could be a valuable tool for both economists and econometricians who are interested in machine learning.

Appendix 1

See Table 6

Table 6 Use of Keywords according to Year

Years	Web of science		Scopus	
	Frequently used keywords	Frequency* (respectively)	Frequently used keywords	Frequency*(respectively)
2010	Data Mining, Economics	4,2	Machine Learning, Data Mining	3,2
2011	Economics, Machine Learning, Algorithm, Big Data, Causality, Econometrics	4,3,2,1,1,1	Machine Learning	8
2012	Data Mining, Economics, Machine Learning	3,2,2	Machine Learning, Big Data	4,2
2013	Data Mining, Economics, Machine Learning, Time Series Analysis	4,3,2,2	Machine Learning, Big Data, Smart Grid, Data Mining, Electronic Health Record, Text Mining, Time Series Analysis	9,6,2,2,2,2,2
2014	Economics, Big Data, Machine Learning, Data Mining, Algorithms, Econometrics	9,7,5,4,4,2,1	Big Data, Machine Learning, Data Mining, Economics, Econometrics, Cloud Computing, Data Analytics, Data Science, Ensemble Learning Healthcare, Modelling, Predictive Analytics	22,11,3,3,3,2,2,2,2,2,2
2015	Big Data, Economics, Machine Learning, Data Mining, Algorithms, Causality-Causal Inference, Econometrics	25,13,8,6,3,2,1	Big Data, Machine Learning, Cloud Computing, Classification, Analytics, Time Series, Data Mining, Finance, Artificial Intelligence, Creative Economy, Data Analytics, Economics, Optimization, Predictive Analytics, Regression, Information Economics, Innovation	66,19,10,4,4,3,3,3,2,2,2,2,2,2,2,2,
2016	Big Data, Economics, Data Mining/Data Analytics, Machine Learning, Time Series Analysis, Causality-Causal Inference, Econometrics, Parameter Estimation	34,22,11,12,16,7,3,4,2	Big Data, Machine Learning, Cloud Computing, Data Mining, Economics, Classification, Econometrics, Artificial Neural Networks, Data Integration, Forecasting, Analytics, Artificial Intelligence, China, Data Management, Data Science, Economic Development, Genetic Algorithm, Optimization, Text Mining, Time Series, Industry 4.0,Internet of Things, Casual Inference, Causality, Data Analysis, Economic Growth, Experimental Economics, Financial Forecasting, Macroeconomics, Neural Network, Regression Analysis	80,25,12,11,8,6,6,4,4,4,3,3,3,3,3,3,3,3,3,3,3,2,2,2 ,2,2,2,2, 2,2,2,2

Table 6 (continued)

Years	Web of science: Frequently used keywords	Web of science: Frequency* (respectively)	Scopus: Frequently used keywords	Scopus: Frequency*(respectively)
2017	Big Data, Economics, Machine Learning, Data Mining/Data Analytic, Artificial Intelligence, Econometrics, Internet of Things, Prediction & Forecasting, Causality-Causal Inference, Behavioural Economics, Computational Social Science, Business Intelligence, Game Theory, Spatial Econometrics, Financial Econometrics	42,28,20,17,5,44,4,3,2,2,2,2, 2,2,1	Big Data, Machine Learning, Internet of Things, Cloud Computing, Classification, Data Mining, Artificial Intelligence, Prediction, Regression, Decision Making, Data Science, Economic Growth, Economics, Analytics, Behavioral Economics, Data Analysis, Deep Learning, Forecasting, Game Theory, Health Economics, Simulation, Spatial Econometrics, Support Vector Machine, Innovation, Time-Series, Behavioral Finance	97,51,10,9,7,7,6,5,4,4,4,4,4,3,3,3,3,3,3,3,3,3,3,2
2018	Big Data, Machine Learning, Economics, Data Mining/Analysis, Prediction & Forecasting Artificial Intelligence, Econometrics, Digital Platforms, Causality-Causal Inference, Deep Learning, Optimization, IOT, Behavioural Economics, Industry 4.0, Healthcare, Spatial Econometrics, Model Selection, Cloud Computing, Explanatory Econometrics, Machine Learning Econometrics	35,35 30,24, 10,10, 10,4, 3,3,3, 3,3,3, 3,3,2, 2,1,1	Big Data, Machine Learning, Cloud Computing, Data Mining, Artificial Intelligence, Industry 4.0, Economics, Naïve Bayes, Internet of Things, Deep Learning, Forecasting, Agriculture, Analytics, Neural Network, Text Mining, Time Series, Behavioral Economics, Causal Inference, Data Analysis, Data Analytics, Economic Growth, Optimization, Sentiment Analysis	98,79,12,12,149,12,5,9,8,6,4,4,4,4,5,3,3,3, 3,3,3,3
2019	Big Data, Machine Learning, Economics, Network Analysis, Artificial Intelligence, Causal Inference /Methods, Deep Learning, Algorithms, Statistical Data Analysis, Social Media, Econometrics, Behavioural Economics, Support Vector Machine, Cloud Computing, Artificial Neural Networks, Classification, Financial Econometrics, Bayesian Econometrics	50,47,38,13,12,9,7,5,5,5,4,4,4,3,3,1	Machine Learning, Big Data, Artificial Intelligence, Deep Learning, Classification, Reinforcement Learning, Neural Networks, Data Science, Economics, Industry 4.0, Forecasting, Data Mining, Time Series, Internet of Things, Data Mining, Prediction, Digital Economy, Sharing Economy, Sustainability, Causal Inference, Cloud Computing, Econometrics, Optimization, Decision Making, Digitalization, Sentiment Analysis, Smart Cities, Support Vector Machine, Text mining, Artificial Neural Network, Behavioral Economics, Data Analytics, Health Economics, Demand Response, Simulation, Cluster Analysis, Innovation	144,128,30,20,13,13,15,10,10,10,9,8,8,7,8,8,7,7, 7,6,6,6,6,5,5,5,10,10,5,4,4,4,4,4,4,8,4

Table 6 (continued)

Years	Web of science – Frequently used keywords	Web of science – Frequency* (respectively)	Scopus – Frequently used keywords	Scopus – Frequency*(respectively)
2020	Economics, Machine Learning, Big Data, Internet ofThings, Prediction, Computational Social Sciences, Deep Learning, Industry 4.0, Neural Networks, Algorithms, Stock Markets, Data Analysis, Data Science, Behavioural Finance, Optimization, Sharing Economy, Support Vector Machine, Natural Language Processing, Econometrics, Financial Econometrics Cloud Computing, Causality-Causal Inference, Structural Econometrics	53,41,36,15,10,9,9, 9,9,6,6,5,5,4,4,3,3,3,4,3,2,1	Machine Learning, Big Data, Artificial Intelligence, Neural Networks, Internet of Things, Covid-19 Pandemic, Artificial Neural Network, Support Vector Machine, Regression, Time Series, Industry 4.0, Behavioral Economics, Deep Learning, Economics, Forecasting, Data Science, Digitalization, Random Forest, Economic Growth, Optimization, Reinforcement Learning, Blockchain, Cloud Computing, China, Clustering, Data Analytics, Data Mining, Digital Economy, Agriculture, Classification, Climate Change, Data Visualization, Decision Making, Econometrics, Financial Econometrics, Genetic Algorithm, Sentiment Analysis, Smart cities, Social Networks, Supervised Learning, Sustainability, Casual Inference, Data Analysis	140,115,35,16,14,13,10,7,10,7,5,21,16,10,9,8,13 ,6,6,6,5,5,5,8,5,5,5,4,4,4,4,4,4,4,4,4,4,4,4,4,3,3

*Frequency implies how many times the specified keyword was used in the specified year

Appendix 2

See Table 7

Table 7 Main congresses that published the highest number of articles and that were ascribed to the most in 2010–2020

Web of science				Scopus			
Number of documents		Number of ascriptions		Number of documents		Number of ascriptions	
Congress	Document	Congress	Citation	Congress	Documents	Congress	Citation
Proceedings of the 21st ACM International Conference on Knowledge Discovery and Data Mining	2	Proceedings of the 2nd International Conference on Open and Big Data (2016)	358	Journal of Physics: Conference Series	52	2nd International Conference on Open and Big Data (2016)	611
Proceedings of the IEEE ACM International Conference on Advances in Social Networks Analysis and Mining (2015)	2	Proceedings of the National Academy of Sciences of the United States	29	ACM International Conference Proceeding Series	50	Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining	194
International Conference on Computational Intelligence (2015)	2	17th IEEE International Conference on Machine Learning	18	CEUR Workshop Proceedings	39	Proceedings of the National Academy of Sciences of the USA	191
IEEE 37th International Conference on Distributed Computing Systems (2017)	2	37th International Conference on Distributed Computing Systems	12	E3S Web of Conference	23	IEEE International Congress on Big Data (2014)	140
IEEE 6th International Congress on Big Data (2017)	2	Proceedings of the Tenth ACM International Conference on Web Search	12	IOP: Conference Series: Earth and Environmental Science	16	IEEE/ACM 6th International Conference on Utility and Cloud Computing (2013)	97
IEEE International Conference on Big Data (2019)	2	Proceedings of the IEEE Second International Conference on Big Data (2016)	11	IOP: Conference Series: Materials Science and Engineering	15	IEEE International Conference on Big Data (2015)	83
13th International Technology, Education and Development Conference	2	Proceedings of the Twelfth ACM International Conference on Web Search	11	Portland International Conference on Management of Engineering and Technology: Managing Technological Entrepreneurship: The Engine for Economic Growth, Proceedings (2018)	12	IEEE International Conference on Services Computing (2011)	76
14th International Technology, Education and Development Conference	2	IEEE 14th International Conference on Industrial Informatics (2016)	9	Proceedings of the National Academy of Sciences of the USA	11	Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining	75
13th International Days of Statistics and Economics	2	Thirtieth AAAI Conference on Artificial Intelligence	7	IEEE International Conference on Big Data (2018)	10		
Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing	1	IEEE International Congress on Big Data (2016)	7	International Conference on Intelligent Transportation, Big Data and Smart City (2015)	8	Proceedings of the Annual ACM Symposium on Theory of Computing	73

Table 7 (continued)

Web of science				Scopus			
Number of documents		Number of ascriptions		Number of documents		Number of ascriptions	
Congress	Document	Congress	Citation	Congress	Documents	Congress	Citation
IEEE 11th International Conference on Ubiquitous Intelligence and Computing (2014)	1	4th InternationalConference on Teaching and Computational Science	7	Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining	7	17th IEEE International Conference on Machine Learning and Applications (2018)	60
Proceedings of the forty-sixth annual ACM Symposium on Theory of Computing	1	6th International Congress on Big Data	6	IEEE International Conference on Big Data (2015)	6	Proceedings of the 23rd International Conference on World Wide Web	45
IEEE Conference on Computational Intelligence for Financial Economics (2014)	1	Third European Network Intelligence Conference (2016)	6	Proceedings of the 28th International Business Information Management Association Conference Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth	6	International Conference on Smart Technologies and Management for Computing Communication, Controls, Energy and Materials (2015)	43
Proceedings IEEE International Conference on Big Data (2015)	1			Ceur Workshop Proceedings	5	Conference on Human Factors in Computing Systems	42
Proceedings of the 29th AAAI Conference on Artificial	1	IEEE 22nd International Conference on Intelligent Engineering (2018)	5	36th International Conference on Machine Learning (2019)	5	25th International Association for Management of Technology Conference Proceedings: Technology – Future Thinking (2016)	38
Proceedings 2nd International Conference on Open and Big Data (2016)	1	25th Australasian Software Engineering Conference (2018)	4	Matec Web of Conferences	5	32nd International Conference on Machine Learning (2015)	37
17th IEEE International Conference on Machine Learning (2018)	1	4th International Scientific Conference: TOSEE	4	Proceedings 16th IEEE International Symposium on Parallel and Distributed Processing with Applications	5	IEEE International Conference on Big Data (2018)	36
Proceedings IEEE Second International Conference on Big Data (2016)	1	Proceedings of the 2016 ACM Conference on Economics and Computation	4	Proceedings of the 29th International Business Information Management Association Conference Education Excellence and Innovation Management Through Vision 2020: From Regional Development Sustainability to Global Economic Growth	5	10th International Conference Management of Large-Scale System Development (2017)	36
IEEE International Congress on Big Data (2016)	1	13th International Technology, Education and Development Conference	3	2nd IEEE International Conference on Cloud Computing and Big Data Analysis (2017)	4	Proceedings of the 12th ACM International Conference on Web Search and Data Mining	30
3rd European Network Intelligence Conference (2016)	1	IEEE International Conference on Big Data (2019)	3	AIES Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics and Society	4	IEEE International Conference on Big Data (2017)	26
				AIP Conference Proceedings		Proceedings of the Conference on Traffic and Transportation Studies	25

References

Aghaei Chadegani A, Salehi H, Md YM, Farhadi H, Fooladi M, Farhadi M, Ale Ebrahim N (2013) A comparison between two main academic literature collections: web of science and scopus databases. Asian Soc Sci 9(5):18–26

Ali L, Alnajjar F, Khan W, Serhani MA, Al Jassmi H (2022) Bibliometric analysis and review of deep learning-based crack detection literature published between 2010 and 2022. Buildings 12(4):432. https://doi.org/10.3390/buildings12040432

Alonso JM, Castiello C, Mencar C (2018) A bibliometric analysis of the explainable artificial intelligence research field. In: Communications in Computer and Information Science Book Series, Vol 853. CCIS

Appio FP, Cesaroni F, Di Minin A (2014) Visualizing the structure and bridges of the intellectual property management and strategy literature: a document co-citation analysis. Scientometrics 101(1):623–661

Ardito L, Scuotto V, Del Giudice M, Petruzzelli AM (2019) A bibliometric analysis of research on big data analytics for business and management. Manag Decis 57(8):1993–2009. https://doi.org/10.1108/MD-07-2018-0754

Athey S (2017) Beyond prediction: using big data for policy problems. Science 355(6324):483–485. https://doi.org/10.1126/science.aal4321

Athey S, Imbens GW (2017) The state of applied econometrics: causality and policy evaluation. J Econ Perspect 31(2):3–32. https://doi.org/10.1257/jep.31.2.3

Athey S (2015) Machine learning and causal inference for policy evaluation. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. ACM pp 5–6

Athey S (2018) the impact of machine learning on economics in "The economics of artificial intelligence: an agenda," National Bureau of Economic Research, Inc, http://www.nber.org/chapters/c14009.pdf

Backhaus K, Lügger K, Koch M (2011) The structure and evolution of business-to- business marketing: a citation and co-citation analysis. Ind Mark Manage 40(6):940–951

Baker HK, Kumar S, Pattnaik D (2021) Twenty-five years of the journal of corporate finance: a scientometric analysis. J Corp Finance 66:101572

Belmonte JL, Segura Robles A, Moreno Guerrero AJ, Parra Gonzalez ME (2020) Machine learning and big data in the impact literature: a bibliometric review with scientific mapping in web of science. Symmetry 12(495):1–15

Bhattacharya S (2019) Some salient aspects of machine learning research: a bibliometric analysis. J Scientometr Res 8(2s):85–92

Bidwe RV, Mishra S, Patil S, Shaw K, Vora DR, Kotecha K, Zope B (2022) Deep learning approaches for video compression: a bibliometric analysis. Big Data Cogn Comput 6:44. https://doi.org/10.3390/bdcc6020044

Bonilla CA, Merigó JM, Torres-Abad C (2015) Economics in Latin America: a bibliometric analysis. Scientometrics 105(2):1239–1252

Bortoluzzi M, de Souza CC, Furlan M (2021) Bibliometric analysis of renewable energy types using key performance indicators and multicriteria decision models. Renew Sustain Energy Rev 143:110958

Castillo-Vergara M, Alvarez-Marin A, Placencio-Hidalgo D (2018) A bibliometric analysis of creativity in the field of business economics. J Bus Res 85:1–9

Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F (2011) Science mapping software tools: review, analysis, and cooperative study among tools. J Am Soc Inform Sci Technol 62:1382–1402

Cortés-Sánchez JD (2020) A bibliometric outlook of the most cited documents in business, management and accounting in Ibero-America, European research on management and business. Economics 26(2020):1–8

Daim T, Ünsal-Rueda GR, Martin HT (2005) Technology forecasting using bibliometric analysis and system dynamics. In: Conference: technology management: a unifying discipline for melting the boundaries. Weber IEEE, USA, pp 112–122

Demirkol D, Önay Koçoğlu F, Aktaş Ş, Erol Ç (2022) A bibliometric analysis of the relationship between diabetes and artificial intelligence. J Istanb Fac Med 85(2):249–257

Dhamija P, Bag S (2020) Role of artificial intelligence in operations environment: a review and bibliometric analysis. TQM J 32(4):869–896. https://doi.org/10.1108/TQM-10-2019-0243

Donthu N, Kumar S, Pandey N, Soni G (2020) A retrospective overview of Asia Pacific journal of marketing and logistics using a bibliometric analysis. Asia Pac J Mark Logist. https://doi.org/10.1108/APJML-04-2020-0216

Donthu N, Kumar S, Mukherjee D, Pandey NB, Lim WM (2021) How to conduct a bibliometric analysis: an overview and guidelines. J Bus Res 133:285–296

Dos Santos BS, Steiner MT, Fenerich AT, Lima RH (2019) Data mining and machine learning techniques applied to public health problems: a bibliometric analysis from 2009 to 2018. J Comput Ind Eng 138:1–11

Durisin B, Puzone F (2009) Maturation of corporate governance research, 1993–2007: an assessment. Corp Gov Int Rev 17(3):266–291

Einav L, Levin J (2014) The data revolution and economic analysis, Technical Report, NBER Innovation Policy and the Economy Conference

Elie L, Granier C, Rigot S (2021) The different types of renewable energy finance: a Bibliometric analysis. Energy Econ 93:104997

Ellegaard O, Wallin JA (2015) The bibliometric analysis of scholarly production: how great is the Impact? Scientometrics 105(3):1809–1831

Fahimnia B, Sarkis J, Davarzani H (2015) Green supply chain management: a review and bibliometric analysis. Int J Prod Econ 162:101–114. https://doi.org/10.1016/j.ijpe.2015.01.003

Fan J, Han F, Liu H (2014) Challenges of Big Data analysis. Natl Sci Rev 1:293–314. https://doi.org/10.1093/nsr/nwt032

Farrukh M, Meng F, Wu Y, Nawaz K (2020) Twenty-eight years of business strategy and the environment research: a bibliometric analysis. Bus Strateg Environ 29(6):2572–2582. https://doi.org/10.1002/bse.2521

Ferreira MP, Pinto CF, Ribeiro SFA, Florindo G (2011) John Dunning's influence in International business/strategy research: a bibliometric study in the strategic management journal. J Strateg Manag Educ 7(2):1–24

Foley AM, Leahy PG, Marvuglia A, Mckeogh EJ, (2012) Renew Energy 37(1):1–8

Fonseca BPF, Sampaio RB, Fonseca MVA, Zicker F (2016) Co-authorship network analysis in health research: method and potential use. Health Res Policy Syst 14:34. https://doi.org/10.1186/s12961-016-0104-5

Gao P, Meng F, Mata MN, Martins JM, Iqbal S, Correia AB, Dantas RM, Waheed A, Xavier Rita J, Farrukh M (2021) Trends and future research in electronic marketing: a bibliometric analysis of twenty years. J Theor Appl Electron Commer Res 16(5):1667–1679. https://doi.org/10.3390/jtaer16050094

Goyal K, Kumar S (2021) Financial literacy: a systematic review and bibliometric analysis. Int J Consum Stud 45(1):80–105. https://doi.org/10.1111/ijcs.12605

Gürsakal N (2009) Sosyal Ağ Analizi, Dora Yayınları

Hjørland B (2013) Facet analysis: the logical approach to knowledge organization. Inf Process Manage 49(2):545–557

Hu C, Song M, Guo F (2019) Intellectual structure of market orientation: a citation/co-citation analysis. Mark Intell Plan 37(6):598–616

Julius R, Halim MSA, Hadi NA, Alias AN, Khalid MHM, Mahfodz Z, Ramli FF (2021) Bibliometric analysis of research in mathematics education using scopus database. EURASIA J Math Sci Technol Educ. https://doi.org/10.29333/ejmste/11329

Kalantari A, Kamsin A, Kamaruddin HS, Ebrahim NE, Gani A, Ebrahimi A, Shamshirband S (2017) A bibliometric approach to tracking big data research trends. J Big Data 4:30

Knoke D, Yang S (2020) Social network analysis, quantitative applications in the social sciences 154, 3rd edn. SAGE Publications Inc, Thousand Oaks

Kumar S, Surekha R, Lim WM, Mangla SK, Goyal N (2021a) What do we know about business strategy and environmental research? Insights from business strategy and the environment. Bus Strategy Environ. https://doi.org/10.1002/bse.2813

Kumar S, Lim WM, Pandey N, Westland JC (2021b) 20 years of electronic commerce research. Electron Commer Res 21(1):1–40

Li Y, Xu Z, Wang X (2020) A bibliometric analysis on deep learning during 2007–2019. Int J Mach Learn Cybern 11:2808–2826

Liao H, Tang M, Luo L, Li C, Chiclana F, Zeng X-J (2018) A bibliometric analysis and visualization of medical big data research. Sustainability 10(12):1–18

Linden R, Barbosa LF, Digiampietri LA (2017) "Brazilian style science"—an analysis of the difference between Brazilian and international Computer Science departments and graduate programs using social networks analysis and bibliometrics. Soc Netw Anal Min 7:44

Linnenluecke MK, Chen X, Ling X, Smith T, Zhu Y (2017) Research in finance: a review of influential publications and a research agenda. Pac Basin Finance J 43(April):188–199

Liu X, Sun R, Wang S, Wu Y (2019) The research landscape of big data: a bibliometric analysis. Library Hi Tech 38(2):367–384

Makawana PR, Jhaveri RH (2017) A bibliometric analysis of recent research on machine learning cyber security. Lecture Notes in Networks and Systems Book Series, LNNS 19

Mao M, Li Z, Zhao Z, Zeng L (2018) Bibliometric analysis of the deep learning research status with the data from web of science. In: Tan Y, Shi Y, Tang Q (eds) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol 10943. Springer, Cham

Merediz-Solà I, Bariviera AF (2019) A bibliometric analysis of bitcoin scientific production. Res Int Bus Finance 50:294–305. https://doi.org/10.1016/j.ribaf.2019.06.008

Merigó JM, Yang JB (2017) A bibliometric analysis of operations research and management science. Omega 73:37–48. https://doi.org/10.1016/j.omega.2016.12.004

Mishra D, Gunasekaran A, Papadopoulos T (2018) Big data and supply chain management: a review and bibliometric analysis. ANN Oper Res 270:313–336

Miskiewicz R (2020) Internet of things in marketing: bibliometric analysis. Mark Manag Innov 3:371–381. https://doi.org/10.21272/mmi.2020.3-27

Nakhodchi S, Dehghantanha A (2020) A bibliometric analysis on the application of deep learning in cybersecurity. In: Karimipour H, Srikantha P, Farag H, Wei-Kocsis J (eds) Security of cyber-physical systems. Springer, Cham

Newman MEJ (2001) Scientific collaboration networks. I. Network construction and fundamental results. Phys Rev E 64:016131

Nicholas D, Ritchie M (1978) Literature and bibliometrics. Clive Bingley, London

Perez-Aranda J, Pelaez-Verdet A (2021) An aplication of social network mining to scientific data: identifyinf networks structures and detecting partnerships in metrics and citation patterns. Soc Netw Anal Min 11:4

Perianes-Rodríguez A, Olmeda-Gómez C, Moya-Anegón F (2010) Detecting, identifying and visualizing research groups in co-authorship networks. Scientometrics 82:307–319

Pesta B, Fuerst J, Kirkegaard OW (2018) Bibliometric keyword analysis across seventeen years (2000–2016). J Intell Artic 6:46

Pieters R, Baumgartner H (2002) Who talks to whom? Intra-and interdisciplinary communication of economics journals. J Econ Lit 40(2):483–509

Rialti R, Marzi G, Ciappei C, Busso D (2019) Big data and dynamic capabilities: a bibliometric analysis and systematic literature review. Discov J 57(8):2052–2068

Rosokhata A, Minchenko M, Khomenko L, Chygryn O (2021) Renewable energy: a bibliometric analysis. In: E3S web of conferences, Vol. 250. EDP Sciences, p 03002)

Salod Z, Singh Y (2020) A five-year (2015 to 2019) analysis of studies focused on breast cancer prediction using machine learning: a systematic review and bibliometric analysis, systematic reviews and meta-analysis. Department of TeleHealth University of KwaZulu-Natal Durban South Africa

Sarkodie SA, Owusu PA (2020) Bibliometric analysis of water–energy–food nexus: sustainability assessment of renewable energy. Curr Opin Environ Sci Health 13:29–34

Scott J (1988) Social network analysis. Sociology 22:1. https://doi.org/10.1177/0038038588022001007

Sengupta N (2015) Machine L techniques in applied econometrics. Carnegie Mellon Tepper School of Business, Doctorate Thesis

Sönmez ÖF (2020) Bibliometric analysis of educational research articles published in the field of social study education based on web of science database. Particip Educ Res 7(2):216–229. https://doi.org/10.17275/per.20.30.7.2

Sthle L, Wold S (1989) Analysis of variance (ANOVA). Chemom Intell Lab Syst 6(4):259–272

Stout NL, Alfano CM, Belter CW, Nitkin R, Cernich A, Siegel KL, Chan L (2018) A bibliometric analysis of the landscape of cancer rehabilitation research (1992–2016). JNCI-J Natl Cancer Inst 110(8):815–824. https://doi.org/10.1093/jnci/djy108

Stremersch S, Verniers I, Verhoef PC (2007) The quest for citations: drivers of article impact. J Mark 71(3):171–193

Su M, Peng H, Li S (2021) A visualized bibliometric analysis of mapping research trends of machine learning in engineering (MLE). Expert Syst Appl 186:115728

Taylor L, Schroeder R, Meyer E (2014) Emerging practices and perspectives on big data analysis in economics: bigger and better or more of the same? Big Data Soc. https://doi.org/10.1177/2053951714536877

Tran B, Vu G, Ha G, Vuong Q-H, Ho M-T, Vuong T-T, La V-P et al (2019) Global evolution of research in artificial intelligence in health and medicine: a bibliometric study. J Clin Med 8(3):360. https://doi.org/10.3390/jcm8030360

van Eck NJ, Waltman L (2017) Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics 111(2):1053–1070

van Eck NJ, Waltman L (2019) VOSviewer Manual. Leiden University, The Netherlands

Waltman L, van Eck NJ, Noyons ECM (2010) A unified approach to mapping and clustering of bibliometric networks. J Informetr 4(4):629–635

Wang K, Feng C, Li M, Pei Q, Li Y, Zhu H, Song X, Pei H, Tan F (2020a) A bibliometric analysis of 23,492 publications on rectal cancer by machine learning: basic medical research is needed. Ther Adv Gastroenterol 13:1–11

Wang X, Xu Z, Škare M (2020b) A bibliometric analysis of economic research-ekonomska istraživanja (2007–2019). Econ Res Ekon Istraž 33(1):865–886. https://doi.org/10.1080/1331677X.2020.1737558

Xian H, Madhavan K (2014) Anatomy of scholarly collaboration in engineering education: a big-data bibliometric analysis. Res J Eng Educ 103(3):486–514. https://doi.org/10.1002/jee.20052

Xu X, Chen X, Jia F, Brown S, Gong Y, Xu Y (2018) Supply chain finance: a systematic literature review and bibliometric analysis. Int J Prod Econ 204:160–173
Yoopetch C, Nimsai S (2019) Science mapping the knowledge base on sustainable tourism development, 1990–2018. Sustainability 11(13):1–17
Zupic I, Cater T (2015) Bibliometric methods in management and organization. Organ Res Methods 18(3):429–472
Azaria A, Ekblaw A, Vieira T, Lippman A (2016) MedRec: Using Blockchain for Medical Data Access and Permission Management, 2nd International Conference on Open and Big Data (OBD), AUG 22-24, 2016, Vienna, Austria.
Ghose A, (2012) Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content, Market SCI, V31, P493, https://doi.org/10.1287/mksc.1110.0700
Huang M-H, Rust RT (2018) Artificial Intelligence in Service, Journal of Service Research, 2018;21(2):155-172. https://doi.org/10.1177/1094670517752459
Mullainathan S (2017) Machine Learning: An Applied Econometric Approach, Journal Economic Perspective, V31, P87, DOI 10.1257/jep.31.2.87
Loebbecke C (2015) Reflections on Societal and Business Model Transformation Arising from Digitization and Big Data Analytics: A Research Agenda, Journal Strategic INF SYST, V24, P149. https://doi.org/10.1016/j.jsis.2015.08.002
Lu R, Zhu H, Liu X, Liu JK, Shao J, (2014) Toward Efficient and Privacy-Preserving Computing in Big Data Era, Publisher: IEEE, V:28, I:4
Varian HR (2014) Big Data: New Tricks for Econometrics, Journal of Economic Perspective, 28 (2): 3-28.
Schneider A, Friedl MA, & Potere D (2010) Mapping Global Urban Areas Using MODIS 500-m Data: New Methods and Datasets Based on 'Urban Ecoregions'. Remote Sensing of Environment, 114(8), 1733-1746.
Zhao JL (2014) Business Challenges and Research Directions of Management Analytics in the Big Data Era, J Manag Analitics, V1, P169, https://doi.org/10.1080/23270012.2014.968643
Zuboff S (2015) Big Other: Surveillance Capitalism and The Prospects of an Information Civilization. Journal of Information Technology, 30(1), 75-89.