UTISGAD
International Journal of Commerce, Industry and Entrepreneurship Studies
ISSN: 2791-6987

Article ID:	UTISGAD-2023-3-2-2807	pp. 67-81
Article Type:	Research Article	
Received:	28.07.2023 *(Abstract)*	
Accepted:	28.12.2023	
Copyright: CC BY-NC 4.0 | Checked by: iThenticate		
Open-Access Policy: BOAI has been applied.		
www.utisgad.org | www.sitso.org.tr | Article Link		

ANALYSIS OF RESEARCH ON MACHINE LEARNING AND MANAGEMENT INFORMATION SYSTEMS USING WEB OF SCIENCE AND R SOFTWARE

İncilay YILDIZ[1]
Fulya ODUNCU[2]

ABSTRACT

The aim of the study is to analyze the studies on Management Information Systems (MIS) and Machine Learning (ML). MIS and ML were questioned by selecting all fields using the Web of Science database for analysis. As a result of the scanning, it was determined that there were 84 academic studies from 2000 to 2023. All studies were included in the bibliometric analysis. 4.3.1 of the R software for analysis. The bibliometric library of the version was used. The overview, sources, authors, documents, conceptual structure and social structure groups in the library were analyzed. The most published years are 2022 (14) 2021 (13) and 2020 (11). The most published journals are ACM Transactıons on MIS (28) and Journal Of MIS (17). The most published author is CHEN H with 6 publications. The countries that publications the most are USA (30) and China (11). The most cited countries are USA (537) and Denmark (214). The most cited study is the article titled "Internet of Things in arable farming: Implementation, applications, challenges and potential" with 116 citations and published in the Biosystems Engineering journal, of which Henriksen is the co-author. It was determined that 216 keywords were used. The most used keywords were determined as Model (10) and Behavior (6).

Keywords: Management Information Systems, Machine Learning, Bibliometric Analysis, Web of Science

JEL Codes: M10, C1, C13

[1] *Assistant Prof. Dr., Altınbaş University, Faculty of Applied Sciences, Management Information Systems, incilay.yildiz@altinbas.edu.tr ORCID: 0000-0001-5572-5058*
[2] *Assistant Prof. Dr., Altınbaş University, Faculty of Applied Sciences, Management Information Systems, fulya.oduncu@altınbas.edu.tr ORCID: 0000-0002-1340-9273*

1. INTRODUCTION

Management, supports decision-making process by identifying decision needs and providing supporting information. Information is the processing of information, which is used in technical, economic and social communication and is the basis of science, in an orderly and appropriate form, through computers and similar electronic devices. System, a systematic operation of people, machines, and methods organized to accomplish specific tasks. The knowledge of information systems is essential for creating competitive firms, managing global companies, and providing valuable products and services to customers. (Sebetci, 2018). MIS was first used in the mid-1960s to prepare periodic reports on accounting, purchasing, inventory, production, sales and payroll MIS, which is the intersection of management systems and information systems in the globalizing world, enables faster decisions to be made by obtaining accurate information. MIS ensures error-free, easy, convenient and fast operation of the entire system, from managers to business professionals, from entrepreneurs to veterans, thanks to the production, management, security, analysis and design of information from data (Laudon & Laudon, 2004).

Computer modeling of the learning process can be defined as machine learning. The most important goal of artificial intelligence research is to create a computer system that can think. Machine learning aims to provide the computer system with the ability to understand and learn on its own. By establishing and testing the machine learning system, the effectiveness and limits of the learning processes are determined. Machine learning is based on the idea that systems can learn from data and enable automated decision-making (with minimal human intervention). It enables data analysis by creating an analytical model. It includes algorithms and processes that enable learning based on databases and data types (Yıldız, 2023).

ML has become an integral part of the MIS field, playing a key role in improving decision-making processes and optimizing organizational operations. In the field of MIS, machine learning algorithms are used for predictive analysis, data classification, and pattern recognition. These advanced analytical tools contribute to the efficiency of information systems by automating tasks, identifying trends, and facilitating more accurate forecasts. Integrating machine learning into MIS fosters a dynamic and adaptable approach to information management, allowing organizations to quickly respond to evolving market conditions and consumer preferences. Publications in MIS and machine learning can help businesses become more competitive, data-driven, efficient and customer-focused. Therefore, it is important for businesses to monitor and implement developments in these areas. Machine learning can be a valuable tool to strengthen data analysis, forecasting and decision support processes within MIS by using their data more effectively, businesses can gain competitive advantage and make more informed decisions. In this study, bibliometric analysis of studies conducted in the field of MIS and ML was examined.

2. LITERATURE REVIEW

Within the scope of the literature research, the 10 most cited studies on Management Information Systems and Machine Learning were examined in table 1.

Table 1: Results of the Researchs

Author(s) and Year	Results of the Research
Andrés Villa-Henriksen, Gareth T.C. Edwards, Liisa A. Pesonen, Ole Green, Claus Aage Grøn Sørensen, 2020	As a result of the article, the role of smart mobile phones and Android devices used in a wide variety of applications due to their usability, connectivity, interoperability, ease of programmability and computing power is mentioned. It is stated that the capabilities of smart mobile devices will increase with the introduction of 5G in the near future. It is suggested that in addition to intelligent management of WSN, it can solve the problems faced by IoT-based solutions with advanced communication technologies.
Hao Hua Sun Yin, Klaus Langenheldt, Mikkel Harlev, Raghava Rao, Mukkamala & Ravi Vatrapu, 2019	The findings of the study suggest that the inherent anonymity level of the Bitcoin Blockchain is not as robust as commonly perceived, and the potential ownership of a Bitcoin address can be more precisely identified. The paper contributes to three key facets: a) it formulates and validates a novel method for enhancing the anonymity of Bitcoin blockchain transactions, b) it offers an initial assessment of the diverse types of assets within the Bitcoin Blockchain ecosystem, and c) it furnishes an array of tools for implementers and regulators, complemented by a prototype implementation of the method that serves as an evaluative instrument
Naveen Kumar, Deepak Venugopal, Lıangfei Qıu,Andsubodha Kumar, 2018	In this study, we propose a novel approach for detecting spam senders by fitting univariate and multivariate distributions to fundamental features derived from user metadata, aiding in distinguishing them from genuine users. Subsequently, we aggregate the probability outputs obtained from individual distributions into a meta-classifier. Empirically, we demonstrate that this approach more effectively generalizes over unseen data compared to typical approaches using standardized features without considering underlying distributions. Therefore, our approach of combining univariate distributions based on user scores with a multivariate distribution unveils a more robust model for detecting spam senders. This model proves to be more powerful in its ability to generalize over unseen data when compared to typical approaches that rely solely on standardized features.
Hilal Atasoy, N. Greenwood, and Jeffrey Scott McCullough, 2019	An Electronic Health Record (EHR) is a digitalized medical record system designed to collect, manage, and share digital health information. The data obtained through this technology plays a crucial role in supporting medical decision-making and operational processes. EHR systems serve as a platform for data analytics,

	offering the potential for long-term improvements in healthcare quality and efficiency. According to study findings, medical practices can be integrated into EHR software, allowing for operationalization. Algorithms embedded within the system can perform tasks such as checking for drug allergies or potential drug interactions. Additionally, treatment guidelines can be incorporated into the EHR, utilizing patient-specific data to prompt healthcare providers with suggestions or alert them to potentially risky interventions.
Wei Dong, Shaoyi Liao, And Zhongju Zhang, 2018	The research results demonstrate the value of social media financial data and further serve as a concept for using such data to complement traditional fraud detection methods.
Langtao Chen, Aaron Baird, And Detmar Straub, 2019	In the study, based on a panel data set obtained from a discussion board focusing on nine chronic diseases from one of the online health platforms, structural social capital was found to be an important antecedent of the exchange of social support found in the online health community and, interestingly, the provision of social support (e.g., helping others) was associated with health literacy. and it was concluded that it had a stronger effect on improving health-related attitudes than receiving social support.
Weiguo Fan, Michael D. Gordon, Praveen Pathak & Praveen Pathak, 2005	This article describes a methodology that uses genetic programming to find new ranking tasks for a Web-based information search function. It was determined that it was carried out for both the temporary function and the routing function in access by making use of the content and structural information in the Web documents used in the discovery process. It is concluded that the performance of retrieving newly found sorting functions for both retrieval functions is better than the performance found with previously known sorting strategies in the literature on information retrieval.
Weifeng Li, Hsinchun Chen, And Jay F. Nunamaker Jr., 2016	In this study, which emphasizes that examining important underground economy sellers has both practical and academic importance in terms of cybercrime forensics and criminology research, AZSecure text mining system has been developed to identify and profile important sellers by using social media analysis. The system identifies sellers using sentiment analysis of customer reviews and profiles sellers using topic modeling of ads. AZSecure The system demonstrated superior performance compared to all other machine learning methods in the identification of advertising topics, sentiment classification of customer reviews, and profiling of the seller. It achieved an average F-measure ranging between 80 percent and 90 percent.
Christopher A. Ramezan, Timothy A. Warner, Aaron E.	This study investigated the impact of varying training sample sizes, ranging from 40 to 10,000, on the performance of six supervised machine-learning algorithms—SVM, RF, k-NN, NEU, LVQ, GBM—in classifying a high-resolution remotely sensed dataset covering a

Maxwell and Bradley S. Price, 2021	large area. While SVM and NEU initially demonstrated lower accuracy with smaller sample sets, their performance consistently improved as the sample sizes increased, suggesting that larger sample sizes are recommended for these algorithms. In contrast, k-NN generally exhibited lower overall accuracy compared to RF, GBM, SVM, and NEU, especially as the size of the training set increased. LVQ emerged as the least effective classifier, displaying notable variations in overall accuracy when trained from different sets of identical size, even when using large training sets.
Yuqian Xu, Mor Armony, Anindya Ghose, 2021	In this article, researchers created a selection model to measure economic data of service quality indicators by working on a data set from one of the best dating sites in the United States. It has been observed that by adding quality indicators, which is one of the text reviews, the probability of predicting patient selections increases significantly. In addition, it also proves that contextual explanation in the prediction results can characterize the quality perception of people using the site better than numerical ratings found on the same service feature.

3. DATASET AND METHODOLOGY

Web of Science (WoS) is a database of scientific information provided by Clarivate Analytics. WoS is a research platform covering academic literature across a variety of disciplines. WoS covers academic literature on a wide range of subjects, including science, technology, social sciences, and the arts. It includes various disciplines from different fields such as natural sciences, social sciences, medicine, engineering, arts and humanities (https://www.webofscience.com/wos/woscc/basic-search). The WoS database includes research analytical tools that provide information on topics such as Indexing and Summarization, Citation Indexes, Conference Proceedings, citation analyses, most cited articles, inter-institutional comparisons, etc. WoS has a powerful search engine that allows users to perform complex queries and narrow the results with various filters. WoS can be an important resource for researchers and scientists and is used to access up-to-date information on academic literature (https://www.webofscience.com).

Bibliometric analysis is a research method used to quantitatively evaluate and measure scientific literature. This type of analysis aims to obtain various information by applying statistical methods on scientific publications, articles, citation data and similar information (De Solla Price, 1963). Bibliometrics is used to evaluate research trends, scholarly interactions, author influence, and developments in specific topics. Within the scope of analysis, Publication Numbers and Trends, Citation Analysis, Author Analysis, Institution Analysis, Journal Analysis, Co-Authorship Analysis are performed. It is used to measure success in the relevant field and to evaluate the impact of researchers, institutions or journals. Bibliometric analysis is often an important source of information

for decision makers on issues such as determining scientific strategies and policies, distributing research funds and improving scientific communication.

ML is used in MIS to optimize business processes and gain competitive advantage by making more effective decisions. It is important to achieve business goals such as Efficiency and Innovation, Improvement in Decision Making Processes, Risk Analysis and Security, Personalized Services and Customer Satisfaction, Automation in Business Processes, Trend Analysis and Future Forecasting, Competitive Advantage. Therefore, The aim of the study is to analyze studies MIS and Machine Learning for bibliometric analysis, MIS and Machine Learning were queried using the Web of Science database. As a result of the scanning, it was determined that there were 84 academic studies from 2000 to 2023. All studies were included in the bibliometric analysis. 4.3.1 of R software for analysis. The bibliometric library of the version was used. The overview, resources, authors, documents, conceptual structure and social structure groups in the library were analyzed. Research data was collected in July 2023.

4. FINDINGS

The data obtained as a result of the bibliometric analysis of studies conducted in the field of MIS and ML are explained with 15 figures. The analysis includes evaluation of studies in this field by years, authors, countries and keywords.

Year	Articles
2022	14
2021	13
2020	11
2000	9
2023	8
2018	7
2002	5
2019	5
2004	4
2005	2

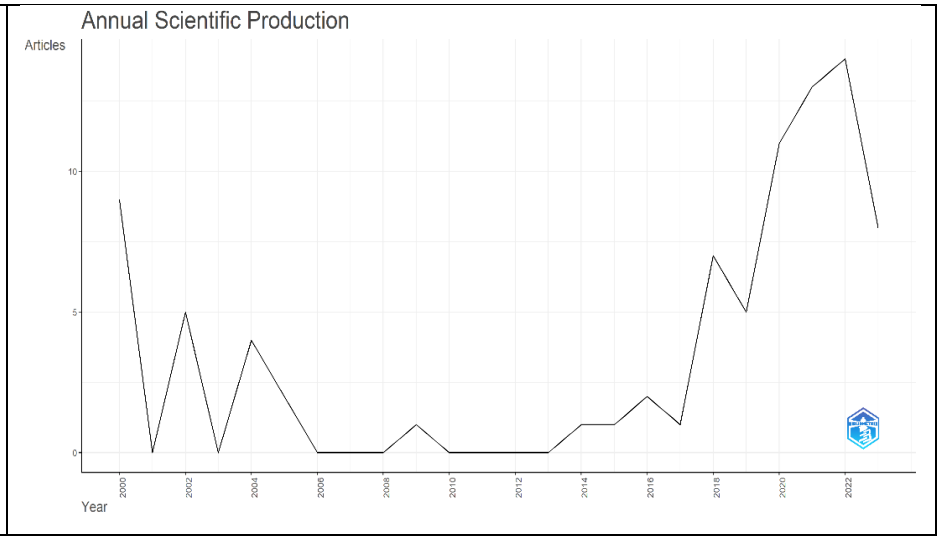

Figure 1. Annual Scientific Production

According to Figure 1, the most studied publication in 2022 was 14 studies. In 2021, 13 studies were carried out. 11 studies were conducted in 2020. It is observed that the number of publications has increased linearly in the last 10 years.

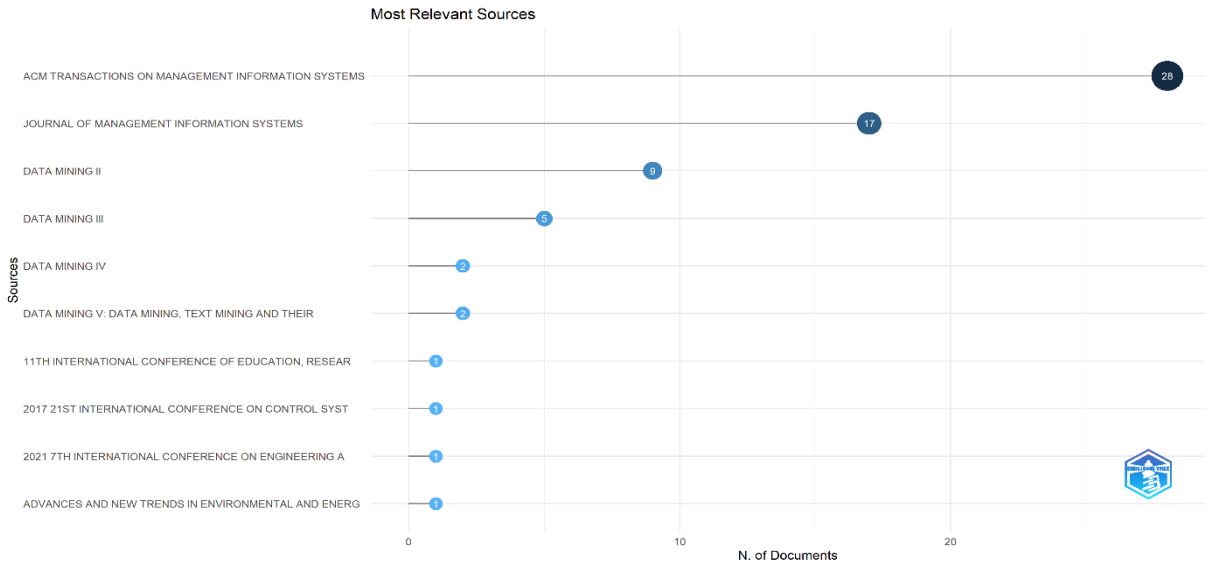

Figure 2. Most Relevant Sources

According to Figure 2, the most published journals are Acm Transactions on MIS (28) and Journal of MIS (17) Data Mining II (9).

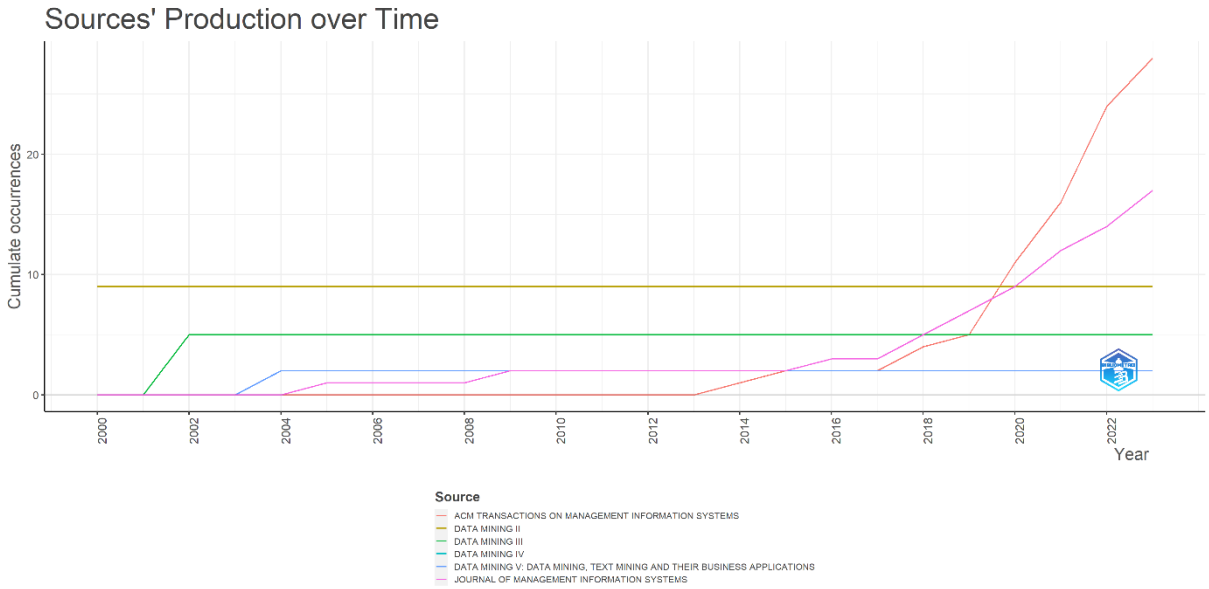

Figure 3. Sources Production Over Time

According to Figure 3, it has been observed that the number of publications of Acm Transactions on MIS has increased after 2012 and has increased rapidly in the last 3 years, ranking first. It is observed that the Journal of Management Information Systems has been active since the 2010s. Data mining magazines appear to be the source with the most publications on this subject. It has been observed that the Data mining II source is the most heavily published data mining source with 9 publications.

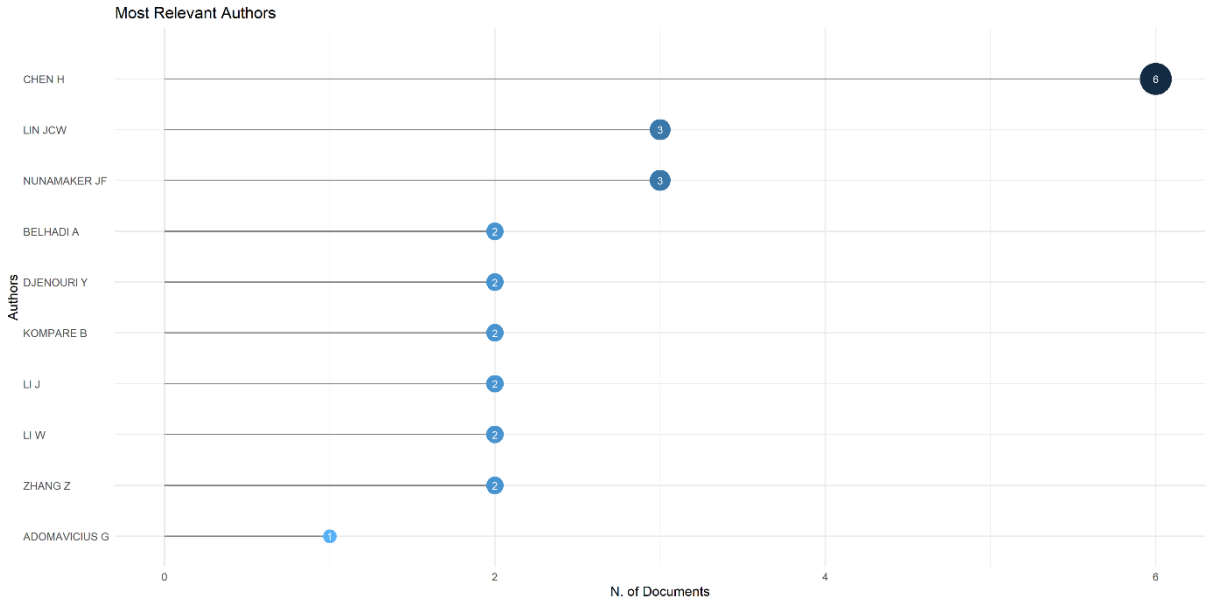

Figure 4. Most Relevant Authors

According to Figure 4, It was determined that the authors named Chen H (9), Lin JCW (3), Nunamaker JF (3) were the authors who published the most on this subject.

Region	Freq
USA	30
CHINA	11
GERMANY	6
AUSTRALIA	3
BRAZIL	3
DENMARK	3
NORWAY	3
FINLAND	2
INDIA	2
SLOVENIA	2

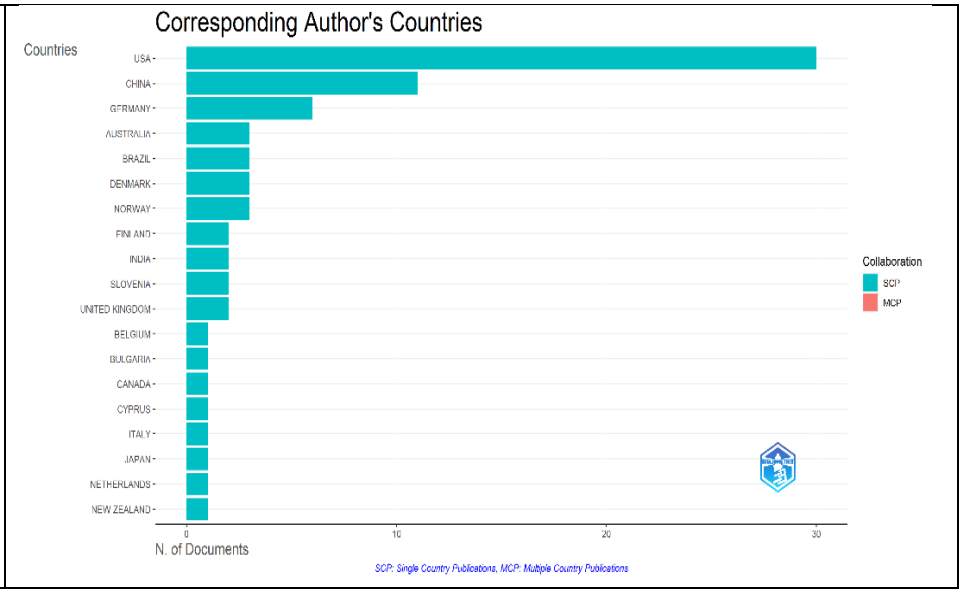

Figure 5. Corresponding Author's Countries

The countries with corresponding authors appear in Figure 5 as the 20 countries with the most publications. The countries with the most documents are the USA (30), China (11), Germany (6). The parts shown in red in the table represent publications made jointly with other countries, and the blue parts represent publications made only with authors from that country. According to Figure 5, it can be seen that there is no joint publication with other countries.

Region	Freq
USA	131
CHINA	46
GERMANY	34
DENMARK	13
NORWAY	13
FINLAND	9
INDIA	8
AUSTRALIA	7
CANADA	7
BRAZIL	6

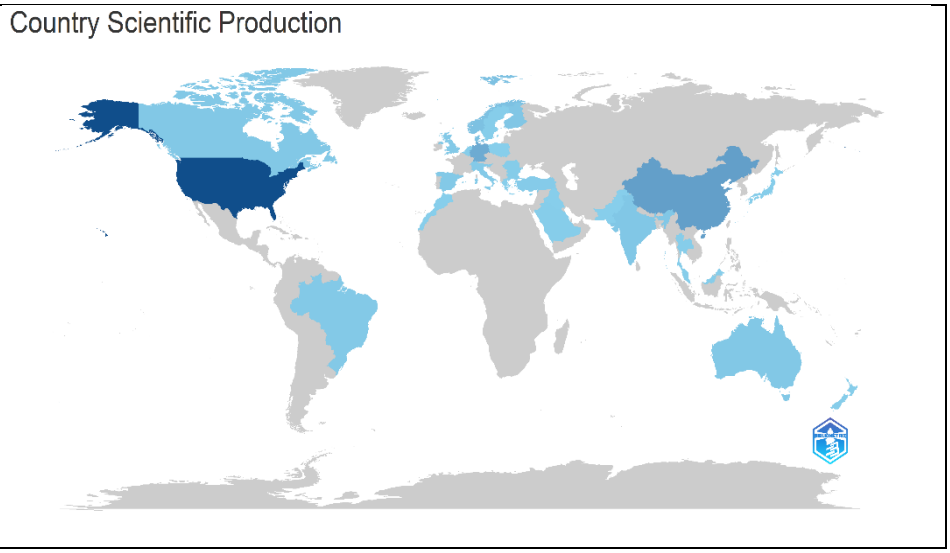

Figure 6. Country Scientific Production

Figure 6 lists the 10 countries with the most publications. Accordingly, the USA ranks first with 131 publications, China ranks 2nd with 46 publications, and Germany ranks 3rd with 34 publications. The countries with the most publications are shown in dark blue. As the number of articles decreases, the blue color turns gray.

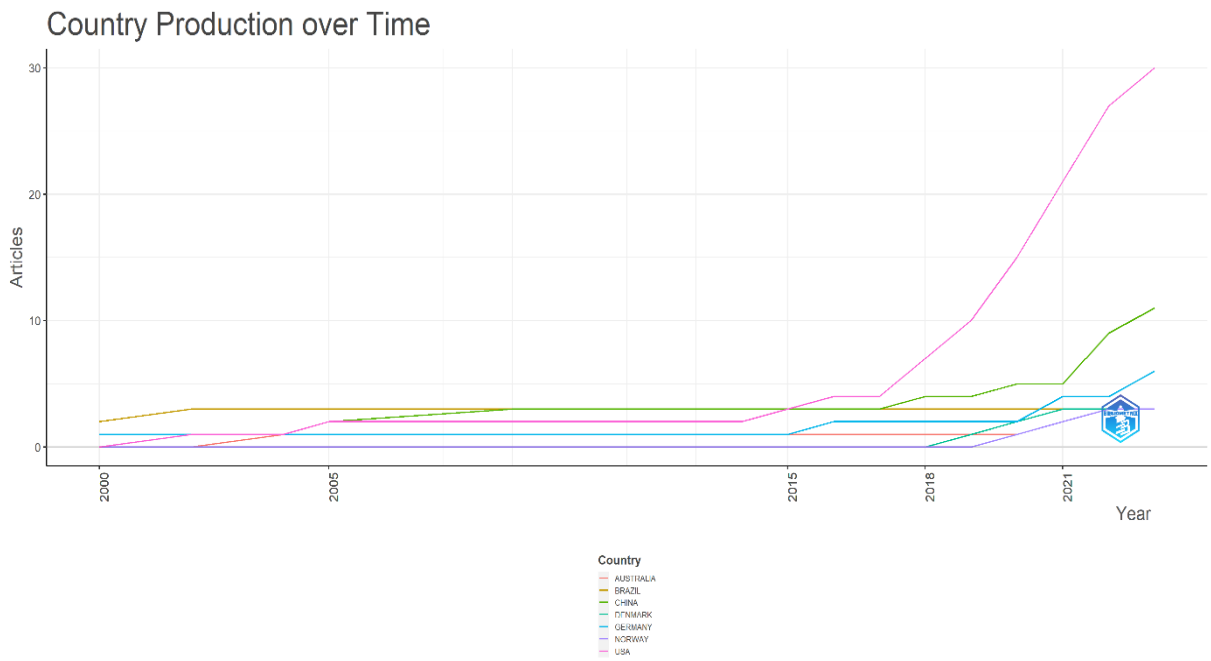

Figure 7. Country Production Over Time

According to Figure 7, a discernible uptrend in publication numbers is evident in the United States after the year 2017. Furthermore, it has been noted that there is a concurrent increase in the number of publications across all countries in recent years.

Country	TC
USA	537
DENMARK	214
CHINA	38
NORWAY	37
GERMANY	19
AUSTRALIA	13
UNITED KINGDOM	12
SPAIN	10
ARAB EMIRATES	9
THAILAND	4

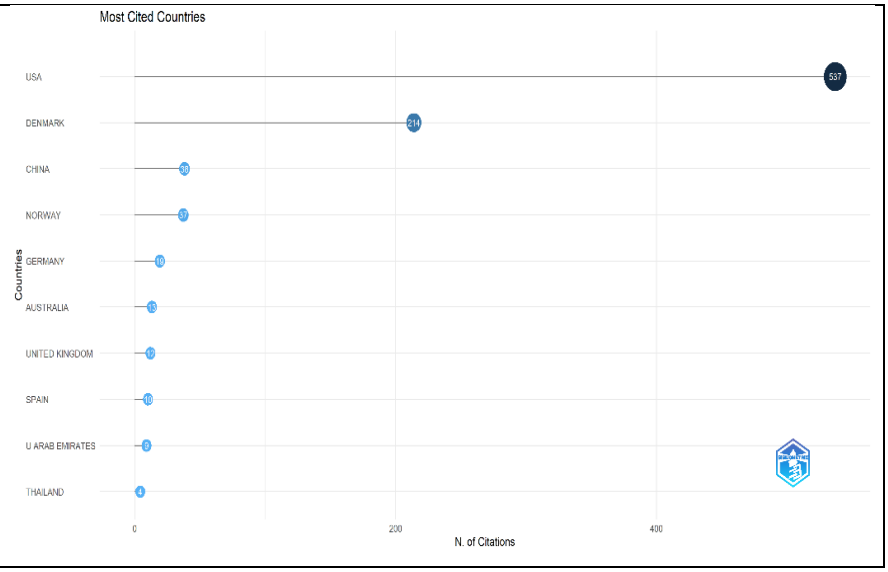

Figure 8. Most Cited Countries

According to Figure 8, the countries of the most cited publications are listed. ABD ranks first with 537, Denmark ranks second with 214, and China ranks third with 38 publications.

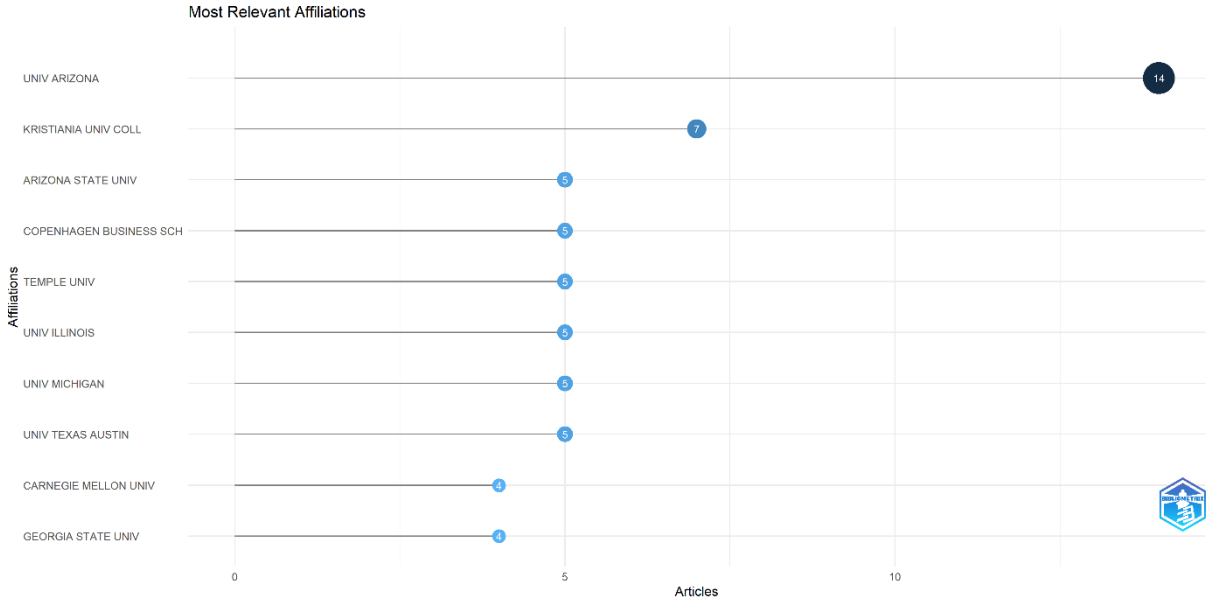

Figure 9. Most Relevant Affilations

Figure 9 lists the institutions that publish the most. University of Arizona and Kristiania University College are at the top with 14 publications and 7 publications.

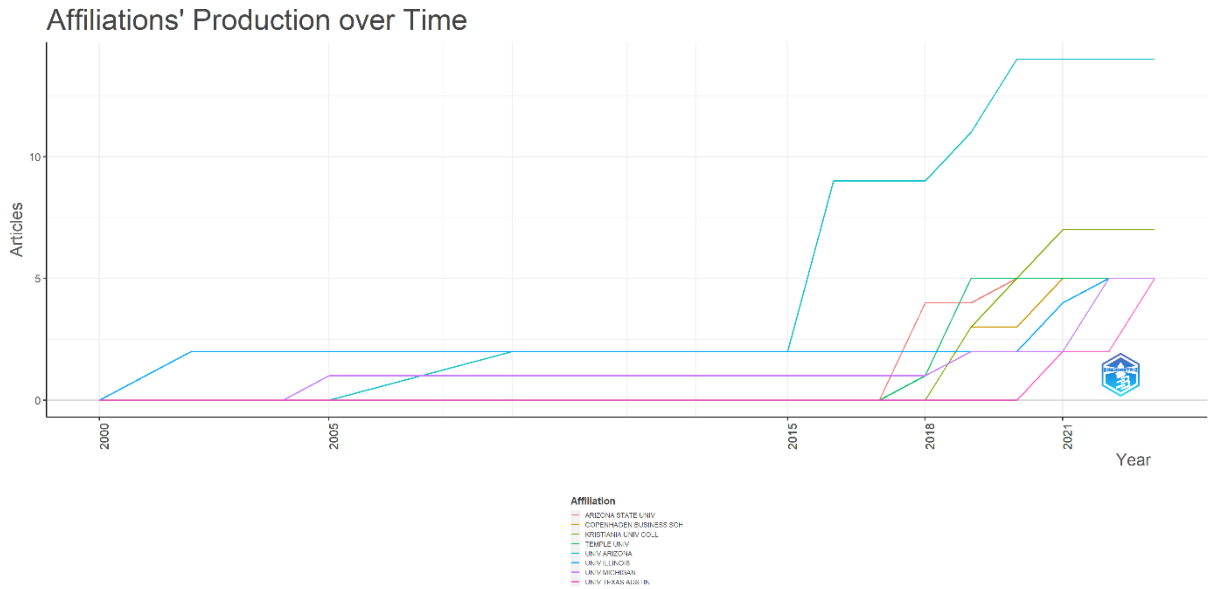

Figure 10. Affilations Production Over Time

Figure 10 shows the change in the number of publications of institutions over the years. According to Figure 10, it is observed that the publications made at the University of Arizona increased in 2015, and the publications made at Kristiania University College increased in 2020.

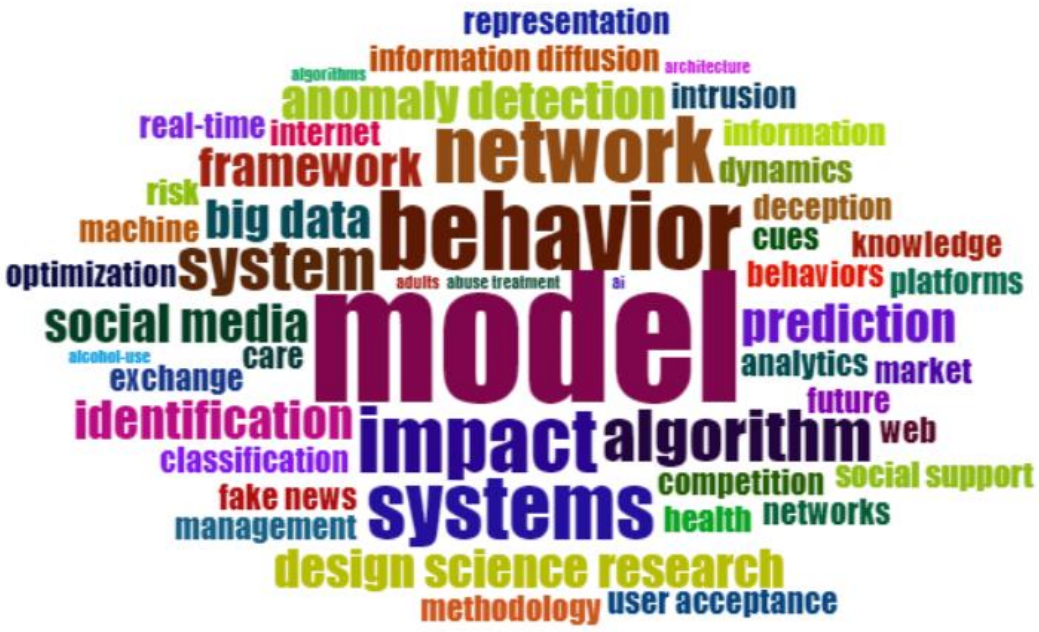

Figure 11. Word Cloud

Figure 11 shows the word cloud with the top 50 most used keywords highlighted. Model and behavior seem to be the most used words.

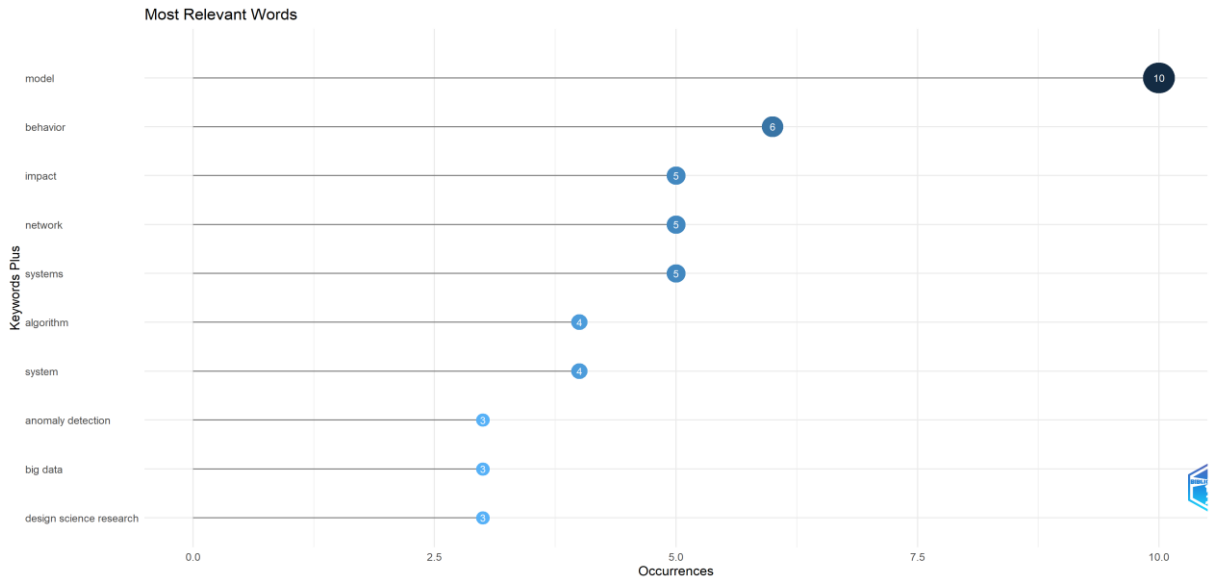

Figure 12. Most Relevant Words

Figure 12 shows the top 10 most used keywords listed. model 10, behavior 8, impact, network, systems were repeated 5 times each.

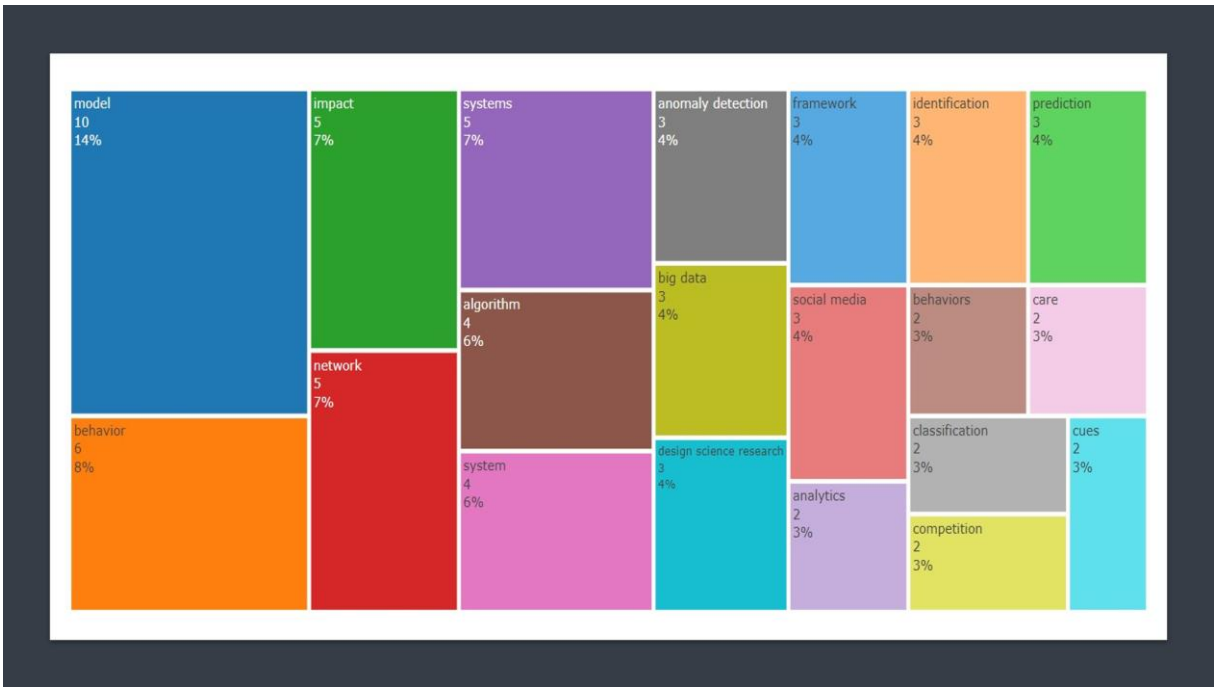

Figure 13. Word Tree

Figure 13 shows the word tree giving the percentages and numbers of the 20 most used words. The importance of the word model was emphasized with 14% for 20 words. It is seen that the importance of prediction and classification models created with machine learning methods in management information systems is revealed.

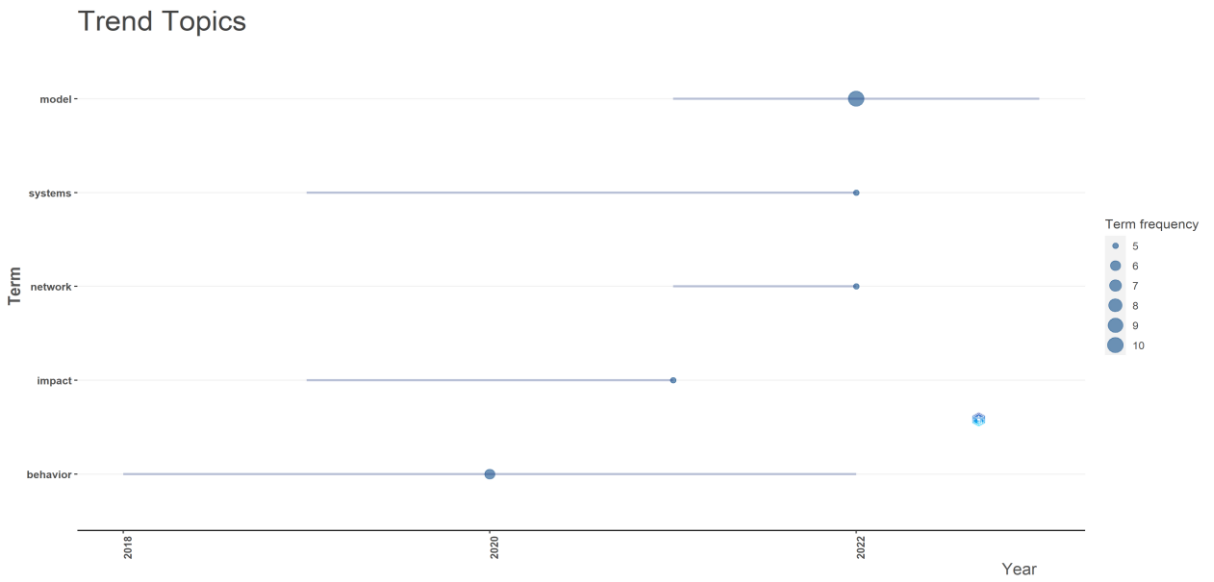

Figure 14. Trend Topics

Figure 14 shows the frequency chart of the first 5 most used words in academic studies according to years. It is seen that the use of the word model has increased in recent years compared to other words.

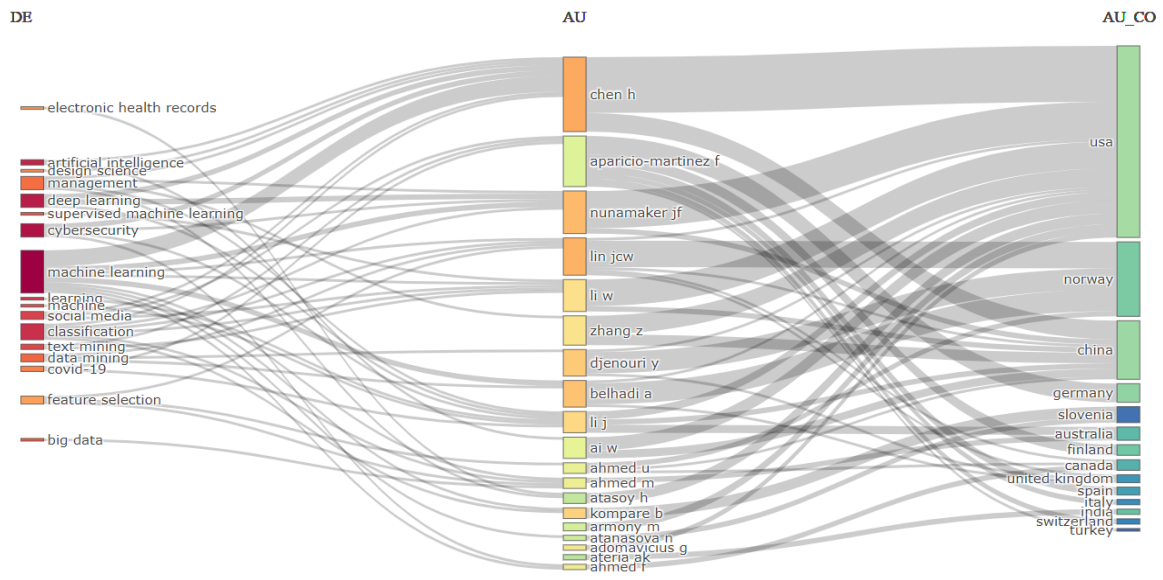

Figure 15. Three-Field Plot

Figure 15 shows the plot containing the keywords, authors and their countries respectively. Figure 15 shows the plot containing the keywords, authors and their countries respectively. We can see the connections with the help of this figure. We can observe the keywords most used by American writers or which country the writers using machine learning are from. This figure provides a summary of the field of study.

5. DISCUSSION AND CONCLUSION

This study is original and important to clarify the place of machine learning processes, which are frequently used in the data field of artificial intelligence (AI) in MIS science. In other studies, AI and MIS studies can be examined at a macro level.

According to the results of the study, 84 publications were identified in 27 sources between 2000 and 2023. While the average number of citations per document was calculated as 10.86, the average age of the documents was 7.18. While it was determined that 290 authors worked in this field, it was observed that only 4 authors worked alone. Co-authorship per document was calculated as 3.63. The international co-authorship percentage is 25 in publications with 4267 references and 216 keywords. The most common document types are articles (55) and academic symposium papers (25). The most published journals are ACM Transactıons on Management Informatıon Systems (28) The most published author is CHEN H with 6 publications. The countries that publications (30) and cites (537) the most are USA. The year with the most publications was 2022. When the findings are examined, it is seen that most publications were made in the last 3 years and that the field is current and important. It is observed that the most cited studies generally emphasize topics such as how machine learning is used in management information systems, its benefits, modeling of algorithms and description of systems.

COMPLIANCE WITH ETHICAL STANDARD

Conflict of Interests: There is no conflict of interest between the authors or any third party individuals or institutions.

Note: This study was presented as a summary paper at the 4th Leadership Academy and Management Congress.

REFERENCES

Atasoy, H., Greenwood, B. N., & McCullough, J. S. (2019). The digitization of patient care: a review of the effects of electronic health records on health care quality and utilization. Annual review of public health, 40, 487-500.

Chen, L., Baird, A., & Straub, D. (2019). Fostering participant health knowledge and attitudes: an econometric study of a chronic disease-focused online health community. Journal of Management Information Systems, 36(1), 194-229.

De Solla Price, D. J. (1963). Little science, big science--and beyond. Columbia University Press.

Dong, W., Liao, S., & Zhang, Z. (2018). Leveraging financial social media data for corporate fraud detection. Journal of Management Information Systems, 35(2), 461-487.

Fan, W., Gordon, M. D., Pathak, P., & PATHAK, P. (2005). Genetic programming-based discovery of ranking functions for effective web search. Journal of Management Information Systems, 21(4), 37-56.

Kumar, N., Venugopal, D., Qiu, L., & Kumar, S. (2018). Detecting review manipulation on online platforms with hierarchical supervised learning. Journal of Management Information Systems, 35(1), 350-380.

Laudon, K. C., & Laudon, J. P. (2004). Management information systems: Managing the digital firm. Pearson Educación.

Li, W., Chen, H., & Nunamaker Jr, J. F. (2016). Identifying and profiling key sellers in cyber carding community: AZSecure text mining system. Journal of Management Information Systems, 33(4), 1059-1086.

Ramezan, C. A., Warner, T. A., Maxwell, A. E., & Price, B. S. (2021). Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data. Remote Sensing, 13(3), 368.

Sebetci, Ö. "Bilişim Sistemleri", Bilgi Teknolojileri ve Yönetim Bilişim Sistemleri, Kodlab Yayın Dağıtım, ISBN: 9786052118221, Ed. Özel Sebetci, 2018

Sun Yin, H. H., Langenheldt, K., Harlev, M., Mukkamala, R. R., & Vatrapu, R. (2019). Regulating cryptocurrencies: a supervised machine learning approach to de-anonymizing the bitcoin blockchain. Journal of Management Information Systems, 36(1), 37-73.

Villa-Henriksen, A., Edwards, G. T., Pesonen, L. A., Green, O., & Sørensen, C. A. G. (2020). Internet of Things in arable farming: Implementation, applications, challenges and potential. Biosystems engineering, 191, 60-84.

Web of science, https://www.webofscience.com/ Avaliable 26.07.2023

Web of science, https://www.webofscience.com/wos/woscc/basic-search Avaliable 26.07.2023

Xu, Y., Armony, M., & Ghose, A. (2021). The interplay between online reviews and physician demand: An empirical investigation. Management Science, 67(12), 7344-7361.

Yıldız, İ. "Zekânın Dijital Dönüşümü; 'Yapay Zekâ'", Dijital Dönüşümün Etkileri, Nobel Bilimsel ISBN: 978-625-398-446-5, 171-194, 2023 Ed. Emir Kaan Cengiz, Fulya Oduncu