European Research Studies Journal
Volume XXIV, Issue 3B, 2021

pp. 1061-1074

Data Science in Decision-Making Processes: A Scientometric Analysis

Submitted 11/07/21, 1st revision 14/08/21, 2nd revision 10/09/21, accepted 30/09/21

Wieslawa Gryncewicz[1], Monika Sitarska-Buba[2]

Abstract:

Purpose: *The article concludes on the importance of scientometric analysis to present research areas and directions in data science in order to support decision-making process.*
Design/Methodology/Approach: *Scientometric analysis.*
Findings: *Article is part of scientometric research performed by authors that results in series of two separate papers. The first one described leading researchers and their area of interest who provide significant input into data science development. The current article quantitatively characterizes the literature thematically related to data science issues, particularly in decision-making processes. The scientometric method was used for data content analysis. The Scopus database was chosen as a source database to perform scientometric analysis. The authors identified core business areas where data science tools have been used in decision-making processes. It is also worth noting the correlation between domain areas and funding sources.*
Practical Implications: *Executing scientific analysis can help to identify research directions in data science area.*
Originality/value: *In our study, we showed that a significant increase in the number of scientific articles in the medical field is directly dependent on research funding institutions. The quantitative characteristics and evolution of keywords, which were the subject of the publications, are also presented. Research directions and their evolution over the years are as well indicated.*

Keywords: *Data science, bibliometric analysis, visualization map, decision-making process.*

JEL Code: *D83, C13.*

Paper type: *Research article.*

Acknowlegdement: *The project is financed by the Ministry of Science and Higher Education in Poland under the programme "Regional Initiative of Excellence" 2019 - 2022 project number 015/RID/2018/19 total funding amount 10 721 040,00 PLN.*

[1]*Ph.D., Wroclaw University of Economics and Business, Wrocław, Poland, e-mail: wieslawa.gryncewicz@ue.wroc.pl;*
[2]*Ph.D. Wroclaw University of Economics and Business, Wrocław, Poland, e-mail: monika.sitarska-buba@ue.wroc.pl;*

1. Introduction

The crucial element in a decision-making process is to have the correct information at the right moment. Nowadays, the problem is not access to information but extracting key pieces for decision making from the 'sea of information' that floods us. IT tools and specialists who know how to use them can help. The ‘volume’ and ‘velocity’ of data generation is so high that traditional database and information system technologies fail to manage and process the data properly. The modern world cannot function without using specialized information and communication technologies because the total amount of data in the world is expected to grow rapidly.

According to the Global Data Protection Index 2020 Snapshot survey released by Dell Technologies, over the past year, the amount of data managed by companies globally has increased by 40 percent. Compared to 2016, this is a jump of 830 percent[3]. The coronavirus pandemic has caused this rate to accelerate further since the first months of 2020. The transition of most companies to remote work, change of the mix of data being created to a richer set of data that includes video communication, and a tangible increase in the consumption of downloaded and streamed video have all contributed to this. The total amount of global data is ever-growing. IDC's Global DataSphere estimates that more than 79 zettabytes (79 trillion gigabytes) of data will be created, captured, copied and consumed globally in 2021, and the amount of data created over the next three years will be more than the data created over the past 30 years[4].

The statistics and forecasts presented clearly show that we are currently experiencing a digital transformation that has significant social and economic consequences. Advanced tools for analysing large data sets are therefore essential. A branch of science called data science (DS), has also emerged. It uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data (Grus 2019). It enables people to make more informed choices than in the past by shaping and filtering the collected data. DS is an interdisciplinary field focused on extracting knowledge from data sets, which are typically large, and applying the knowledge and actionable insights from data to decision-making processes in a wide range of domains.

In the literature, we can find publications and guidebooks on the use of DS in such areas as finance, management, marketing, production, logistics, health care, energy, public administration, social media, and many others (Nymand-Andersen 2021 – in economics and finance; Consoli *et al.*, 2019 – in healthcare; Caviggioli and Ughetto 2019 – in manufacturing, Du and Khan 2020 in information science; Fávero and

[3] *Global Data Protection Index 2020 Snapshot, https://www.delltechnologies.com/en-us/data-protection/gdpi/index.htm*

[4] *Global DataSphere, https://www.idc.com/getdoc.jsp?containerId=IDC_P38353*

Belfiore 2019 in business, Pretty 2013 – in public sector etc). All of these publications address the issues of data analysis and decision process improvement using advances in DS in these particular application areas. There are also a great deal of theoretical scientific works on conceptual frameworks, advanced DS methods and algorithms Menzies *et al.* (2016) present the best practices of seasoned data miners in software engineering. Rao and Rao (2020) highlight principles and new advances in this field. Thwel and Sinha (2020) show various methods that can be used to eliminate duplicated data.

The authors have prepared a series of two articles on the subject. The goal of the first was to identify key institutions and authors working on DS area. The purpose of this paper is to analyse the DS literature using scientometric indicators to identify the research topics that these publications have addressed and how they have changed over the years. We have shown that the increase in the number of scientific articles in most fields is directly dependent on the research funding institutions.

The remainder of the paper is structured as follows: Section two provides a short introduction to data science role in decision-making processes. Afterwards, we make a brief review of related work on scientometric analysis of DS publications on decision-making processes. We also shortly remind the research method and process we used, which were described in detail in the first article of this series. We emphasize that DS tools, techniques and algorithms have been used in various application domains. Section three reports subject area analysis, followed by keyword analysis in section fourth. The keyword analysis focuses on determining what kinds of decisions and processes are supported by the DS framework.

According to the authors, DS is mainly oriented towards supporting decision-making processes through the use and implementation of knowledge derived from data. This section also indicates research directions and their evolution over the years. The article concludes that DS is no longer limited to technology, mathematics, and statistics. Researchers are adopting DS methods, algorithms, and techniques to use their utility to study almost all other aspects of human life. The authors also point out that the development of a discipline is highly dependent on research funding.

2. Related Work and Research Method

A scientometric analysis was chosen as a research method. Scientometric provides a set of tools that allows for a quantitative assessment of a scientific output, deals with the measurement of the impact of research papers and academic journals and statistical evaluation of scientific citations. According to Julia Lane (2009), scientometric methods can assess the dynamics of literature growth in a research area, measure the impact of a scientist's work in a research area, and are used to evaluate the impact of funding on scientific development. There is much more additional features that scientometric offers but these above were visualized in the research paper.

Our study focuses on understanding how research oriented toward application of DS in decision-making processes and related areas has evolved. For this purpose, we also reviewed other works on scientometric analysis of publications in this field. While most of them were based on the Scopus and Web of Science databases, the documents taken for analysis were dated 1983 to 2020 and the VOSviewer software were used for data analysis.

Purnomo *et al.* (2020) researched 5,202 documents published from 1983 to 2019 obtained from the Scopus database. Most were conference papers from the USA and addressed computer science. Ten patterns of collaborative research groups were also noted. Sarkar and Pal (2019; 2020) also focused on various aspects of authorship pattern in DS from 2001 to 2018. In addition, they investigated the annual article growth rate, author productivity rate, and collaboration networks. The results of these analyses were visualised using VOSviewer and GunnMap software. Naturally, among the published papers, those in English language predominated. However, an interesting observation is that they were followed by languages such as German, Chinese, Spanish, French, Japanese, Portuguese, Russian, Polish, and Italian.

In addition to scientometric publications on DS in general, we also encountered many dedicated to the use of DS algorithms to support decision making in different application domains. Tran *et al.* (2019) aimed to evaluate the global development of scientific publications in medicine. Their research has shown interest in such methods in the prognosis and diagnosis of diseases in clinical settings, as well as in robotic-assisted surgery, rehabilitation, and individual data collection for precision medicine. The use of DS approaches in ethical and legal issues is also important, especially regarding privacy and confidentiality of medical data.

Camacho *et al.* (2020) provided a scientometric study to detect the most active research areas and application domains in social network analysis. Perron *et al.* (2017) used DS to create a set of research tools for a scientometric study of social work literature. They observed a rapid and consistent increase in the number of journals in the field, especially from 1997 to 2005.

Another interesting article analyses how the concept of sustainability is being incorporated into global research of higher education (Alejandro-Cruz *et al.*, 2019). This study utilizes different scientometric reviews of global research between 1991 and 2018 using text mining techniques and social network analysis. Based on the results, it was possible to create a list of major institutions that have significantly contributed to the topic of sustainability in higher education, and it was found that 40.58% of the records came from institutions in the United States, China, the United Kingdom and Australia. The study also provides insight into emerging thematic trends and patterns of sustainability research around the world.

Also Xie (2019) deliberates the problematic of DS applications in education, especially in remote education and online learning. He applies the theories and

methods in scientometrics to attraction assessment of massive open online courses. This publication was written 2 years ago. It would be interesting to continue this research after the period when, due to the pandemic, the whole world has basically switched to remote working and learning.

As outlined in the examples above, there are many research papers on scientific mapping of publications addressing DS in various fields. In the next section, we will conduct our own scientometric analysis to demonstrate the impact of DS methods on a decision-making process. As it was already mentioned we are presenting results of our research in two separate articles however, to make our research more comparable we have used the same dataset and method (Gryncewicz and Sitarska-Buba 2021). There are different search criteria available in Scopus database. After evaluating the quality of the content we received based on various filtering criteria, we found the following to be the most effective:

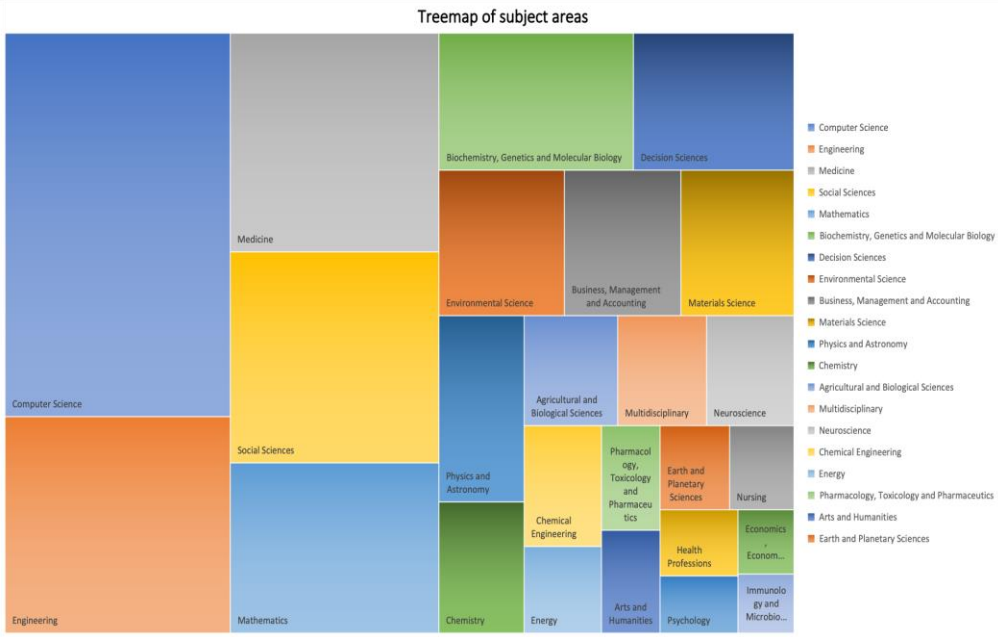
1. 'Data science' string was selected to search within title, abstract and keywords.
2. Search criteria were limited to research articles only.
3. We selected only articles written in English.
4. Finally, dataset have been limited to 'exact keyword=data science'.

Using following criteria we have selected 1685 research articles from Scopus database to build our dataset for scientometric analysis of subject areas and keywords to discover what kind of decision-making processes are supported by data science. To run scientometric analysis the authors used basic data analysis available in Scopus database like subject area analysis but also VOSviewer tool which is a software tool for creating maps based on network data and for visualizing and exploring these maps (Van Eck, 2020). Visualizing maps were created to discover connections between specific keywords to discover decision-making areas where data science tools and technics were used.

3. Subject Areas Analysis

For the initial data set analysis we used scientometric tools that are offered by Scopus itself to create treemap of subject areas. Analysing subject areas of DS statistics, we observe that computer science (18,2%) and engineering (10,3%) are still a significant part of all publications in recent years. However, the applicability of DS methods and technology used in decision support processes is increasing and seems to have become prevalent in the domains of medicine (9,6%), social science (9,3%), biochemistry, genetics and molecular biology (5,6%) and decision sciences (4,6%). DS is no longer limited to technology, mathematics and statistics. Researchers are adopting DS methods, algorithms and techniques to leverage their usability to explore almost all other aspects of human life, including agricultural and biological science, neuroscience, arts, or earth and planetary science (Figure 1).

Figure 1. *Contribution by subject area (treemap)*



Note: *Some articles were included in more than one area, therefore the total number of articles in this figure exceeds the total number of articles included in the final dataset.*
Source: *Own treere map created based on the Scopus data*.

The authors compared and analysed how subject areas have been changed over years (2006-2021). DS term has been established to explain how data can be used in decision processes and leverage data-driven approach. Thanks to rapid technology development and data availability from almost all business domains, statistics and mathematics algorithms were easily implemented to explore a large scope of business domains to highlight the applicability of DS.

Moreover, rapid technology development established and simplified the maintenance, management, and storage of large volume of data. In the author's opinion, both these factors contributed to the final direction and development of DS area. Artificial intelligence, neural networks, decision trees and other algorithms have been an integral part of decision support systems since the beginning, however the usage of them was limited because of cost and technology entry barrier. Thanks to open-source platforms like Anaconda, where built-in programming frameworks like Jupiter Notebooks or R Studio are easy to use, implementation of DS has become significantly easier than before.

In Table 1 the authors present how subject area of DS has been developing form the beginning until now. Computer science has the largest share of all research papers over the years, between 17% and 23%. If we include engineering with an average 8,66% of the share, we will see that only around 30% of all research papers have

been published in pure IT subject area. The rest, around 70% of all publications, has been classified into non-IT subject areas as social science (8%-11%), medicine (6%-12%), or decision science, biochemistry, genetics, and molecular biology or business, management and accounting. That would only confirm the conclusion that DS is mainly focusing on implementation and data-driven decision processes.

Table 1. *Top 10 subject areas: Comparison through the years*

Subject area	2006-2016	Subject area	2017-2019	Subject area	2020-2021(Q1)
Computer	23%	Computer Science	17%	Computer Science	18%
Social Sciences	14%	Medicine	12%	Engineering	11%
Mathematics	11%	Engineering	10%	Medicine	9%
Engineering	9%	Social Sciences	9%	Social Sciences	8%
Decision	6%	Mathematics	7%	Mathematics	7%
Medicine	6%	Biochemistry, Genetics and	7%	Biochemistry, Genetics and Molecular Biology	5%
Biochemistry, Genetics and	5%	Decision Sciences	4%	Environmental Science	5%
Physics and Astronomy	4%	Business, Management and	4%	Materials Science	4%
Business, Management and	3%	Materials Science	3%	Physics and Astronomy	4%
Multidisciplinary	3%	Environmental	3%	Decision Sciences	4%

Source: *Own study.*

Overall pandemic situation forces the research effort to turn to medicine and health care, to build models and tools that help in pandemic decision processes. The share of medicine was increasing constantly from 6% in years 2006-2016, reaching 12% in years 2017-2019. Data set was generated in May 2021 so we can expect that medicine share will increase till end of the year, but even now it is significant and represents 9% of all papers.

The same trend is observed for biochemistry, genetics and molecular biology area. The growth is not so rapid, however 5% achieved in Q1 2021 leads to an assumption that an increasing trend will be established.

The descriptive statistics about bibliographic information related to funding sponsors obtained from Scopus additionally confirm author's hypothesis about significant increase of DS application and usage in domains other than computer science and engineering, especially in medicine. The second and third place in funding sponsor ranking is taken by National Institutes of Health (129 papers) and U.S. Department of Health and Human Services (119 papers) (Table 2).

Table 2. *Top 10 funding sponsor by contribution to the field (N = 1685)*

Funding Sponsor	Documents	[%]
National Science Foundation	151	8,7%
National Institutes of Health	129	7,4%
U.S. Department of Health and Human Services	119	6,9%
European Commission	91	5,2%
National Natural Science Foundation of China	53	3,1%

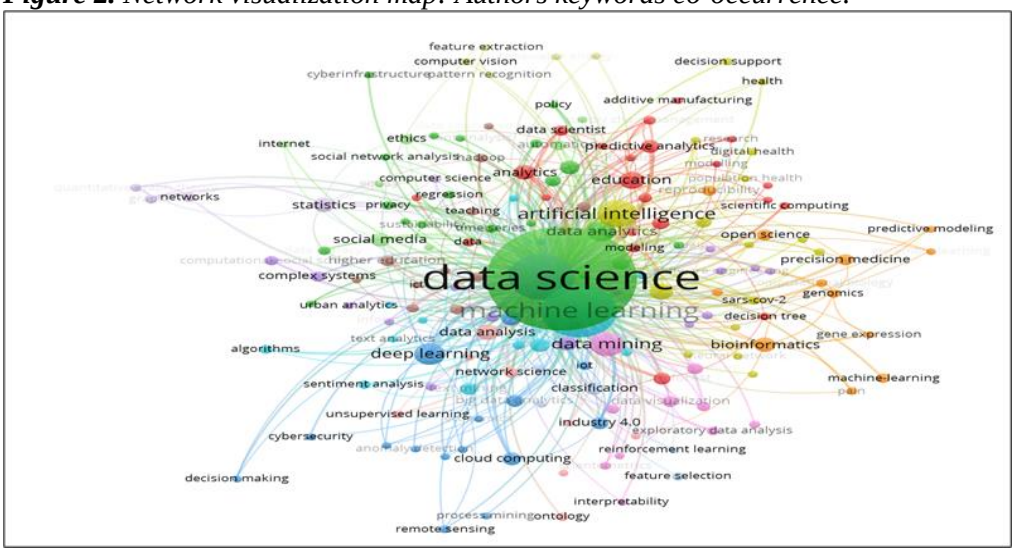
UK Research and Innovation	41	2,4%
Deutsche Forschungsgemeinschaft	40	2,3%
Horizon 2020 Framework Programme	37	2,1%
Engineering and Physical Sciences Research Council	36	2,1%
U.S. Department of Energy	33	1,9%

Note: *Some articles were included in more than one area, therefore the total number of articles in this figure exceeds the total number of articles included in the final dataset.*
Source: *Own data presented based on the Scopus dataset.*

3.1 Keyword's Analysis

We preformed co-occurrence analysis in VOSviewer by selecting author keywords as an analysis unit. The full counting method was chosen. From 4881 keywords we selected 173 because their occurrence was at least 5. As illustrated in Figure 2, indexed keywords indicate that the focal areas of the extant research primarily pertain to DS, machine learning and big data. These findings suggest that only deeper analysis of links and their total strength will allow to identify more precise research areas and how DS algorithms support decision process. Authors analysed all keywords whose Total Link Strength was calculated by VOSViewer up to 6 points.

Figure 2. *Network visualization map: Authors keywords co-occurrence.*



Source: *Visualization map created based on own data created in VOSViewer.*

Authors keywords present several leading areas where DS research is developing towards:

1. *engineering related* (networks, algorithms, natural language processing, machine learning, artificial intelligence, data mining, big data, regression, supervised and unsupervised learning) – there are a lot of publications which focus on improvement, optimization and development of DS algorithms to

build predictive models. Furthermore, keywords such as data visualisation, knowledge discovery, clustering, prediction, and automation, indicate that researchers focus on building algorithms to forecast trends and future behaviours of the population surveyed in the study.

2. *security related* (cyber-infrastructure, cybersecurity, anomaly detection, remote sensing). Security area is one of the leading areas from ongoing digital transformation that has significantly sped up over the last two years. Therefore, it is understandable that one of the directions is the implementation of DS algorithms to predict potential anomaly or to build data driven decision processes for cybersecurity.
3. *medicine related* (precision medicine, electronic health record, digital health, bioinformatics, medical information, pain, and covid-19, gene expression, population health). A large number of medical data actively contributes in data-driven decision processes for diagnosis or forecasting of health behaviour. Keywords such as structural health monitoring or covid-19 indicate an emerging research interest in DS's potential application in advanced health related fields where artificial intelligence and prediction algorithms are used to estimate pandemic development or support medical decisions related to public health.
4. *education related* (teaching, education, higher education, curriculum, data scientist). Increasing popularity of DS as a study object caused researchers to define a curriculum of data scientist training, and also established a list of skills required to become a certified data scientist.

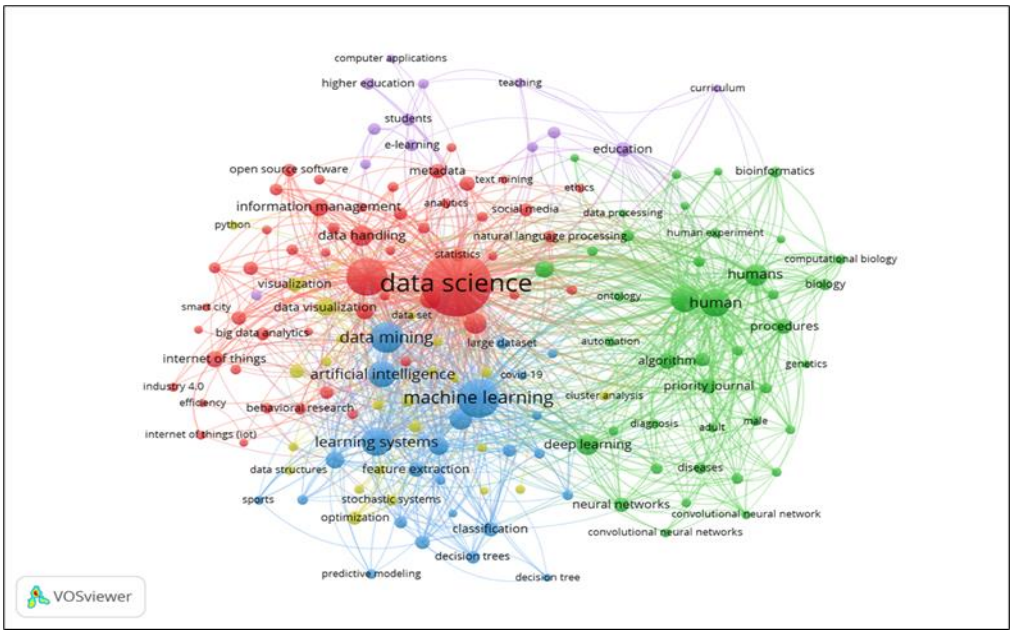
There seems to be increasing interest in potential application of DS in medicine and health related areas and suggested significant academic community's interest in exploring deep learning, machine learning, data and text mining, natural language processing, and analytics implemented in decision support processes for security and education.

Through a deeper keywords analysis authors would like to verify the thesis that computer science and engineering have leading position in subject areas because of scientific classification of data science itself not because content of the papers. We have built four separate datasets limited to subject areas: computer science (605 articles), engineering (341 articles), medicine (319 articles) and social science (308 articles). For each of the dataset we preformed co-occurrence analysis in VOSviewer by selecting author keywords as an analysis unit. The full counting method was chosen.

For *computer science* we selected 155 keywords from 8423 which occurrence was at least 10. The goal was to identify most common keywords therefore the occurrence metric was at higher level than for the whole dataset (1685 papers). Keywords were grouped in 5 clusters, where leading role play technology related wordings like cloud computing, computation theory, cyber physical systems, distributed computer systems, big data analytics, data privacy, data integration, data reduction, learning

systems, artificial intelligence. Next group is also computer science related but focused on data science methods like predictions, cluster analysis, features extractions, optimisation, on algoritms like convolutional neural network or on data visualisation. Only two clusters are showing implementation areas of data science like higher education, computational biology, bioinformatics, smart city, IoT or industry 4.0. Network visualization map is presented on the Figure 3.

Figure 3. *Network visualization map. Authors keywords co-occurrence – computer science dataset*

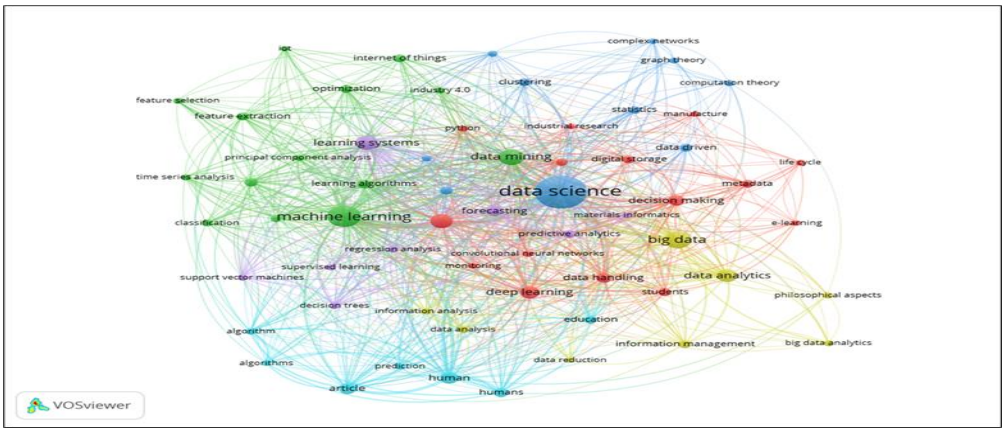


Source: *Visualization map created based on own data created in VOSViewer.*

Network visualisation map created for *engineering* dataset consists of 62 keywords that have been selected from 5743 (Figure 4). The occurrence condition was also set up for 10. Articles that have been classified as engineering are focused on data science learning models like supervised and unsupervised learning, vector machines, classification's algorithms like decision trees, regression, graph theory used in learning processes or time series analysis. Only human or humans as keywords are presenting implementation areas which are not obvious.

Visualisation maps created for computer science and engineering datasets doesn't confirm authors theory that data science classification itself causes so significant share in the subject areas of data science. From the above analysis it can be concluded that DS is still an innovative field, developing and improving its tools at the level of new models, techniques, and algorithms, which is still the subject of scientific research.

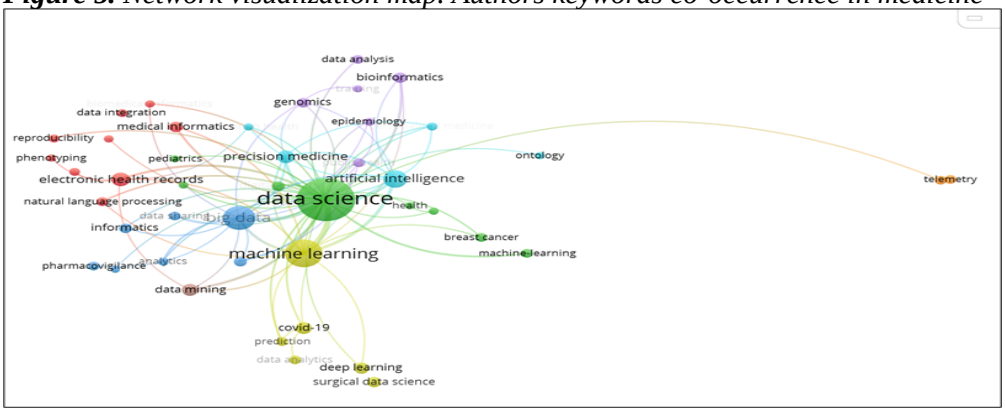
Figure 4. *Network visualization map. Authors keywords co-occurrence- engineering dataset*



Source: *Visualization map created based on own data created in VOSViewer.*

For further keywords analysis, the authors selected medicine to process a more detailed investigation. Medicine's share in DS has been growing rapidly in the last few years, moreover the authors would like also to verify how the ongoing worldwide pandemic influences research directions. 319 articles were selected to run deeper keywords analysis. Extract obtained from Scopus has been limited to Medicine as subject area. VOSviewer identified 1671 authors keywords, only 44 met the threshold that was set up to at least 5 occurrences. Network visualisation map is presented in Figure 5.

Figure 5. *Network visualization map. Authors keywords co-occurrence in medicine*



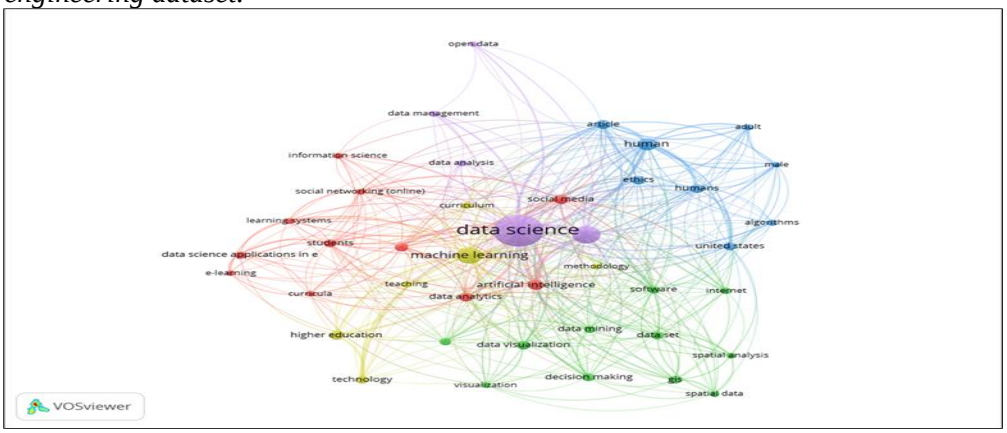
Source: *Visualization map created based on own data created in VOSViewer.*

As expected, the strongest links have been calculated for DS, machine learning and big data, however network visualisation map indicates implementation areas, such as creation of electronic health record, that will help to build models that could be used in precision medicine. Machine learning is strongly connected to COVID-19, prediction keywords, data analytics, deep learning, and surgical DS. Pandemic

predictions and epidemiology are new research subjects that have not been explored before. It is a good example how researchers adopt well know mechanisms to support health related decision processes and public health models. Adoption of DS in medicine is a very promising development direction. A large number of data produced by medicine allows to implement predictive models based on artificial intelligence that will support diagnostic processes and maybe help or even replace initial medical interviews with patients.

For social science we selected 41 keywords from 3271 which occurrence was at least 10 (Figure 6). Network visualization map has built 5 clusters which presets decision making areas where data science can be implemented. In the first cluster implementation areas are as follow e-learning, education, social media, and social networking. The second cluster presents how data science technics are used to present GIS and spatial data. Another grouping is showing strong connection between machine learning used for higher education or curriculum data.

Figure 6. *Network visualization map. Authors keywords co-occurrence – engineering dataset.*



Source: *Visualization map created based on own data created in VOSViewer.*

Medicine and social science keywords analysis is showing that more and more researchers are using data science to improve decision-driven processes by building models that will predict, estimate, and assess different aspects of a business areas.

4. Conclusion

As mentioned in the introduction, this paper is a continuation of a series devoted to the scientometric analysis of scientific papers related to DS. The first paper identified leading researchers and their area of interest, along with having reported statistics that describe top countries and institutions. In this one, based on the collected results, an analysis of the subject areas and keywords that appeared in their publications was conducted. Scientometric indicators were chosen because they

provide multiple techniques for evaluating the content of a scientific publication, taking into account different criteria. They also allow for independent evaluation and the avoidance of subjective bias. In authors opinion visualization maps used during the research have contributed to the connection's discovery that was not originally defined and bring us to unexpected conclusions.

Our research has shown that DS is a scientific discipline which has implementation focus on decision support processes in most publications. Researchers pay attention to building algorithms to predict trends and future behavior to improve decision-making in various application domains. As a result, with the help of DS, managers, regardless of the industry they are in, can make data-driven decisions and make far more precise choices than ever before.

Treemap used to present subject areas analysis showed that computer science and engineering still have a dominant role, what is caused by the scientific classification of data science discipline. Examining deeper the subject areas, we can see that DS is no longer limited to technology, mathematics and statistics. Scientists are adopting DS methods, algorithms and techniques to use their utility to study almost all other aspects of human life, including agricultural and biological sciences, neuroscience, art, or earth and planetary sciences. It was also possible to compare how subject areas have changed over the years (2006-2021). We can observe increasing trend for medicine that achieved in Q1 2021 around 9% of all publication. If medicine would keep the current development trend, we can assume that medicine will be at the first or second place at the end of 2021. Descriptive statistics on bibliographic information related to research funding sponsors also confirmed the authors' hypothesis of a significant increase in interest and use of DS in medicine. In the example we analysed, we can see that in recent years, publications in the field of medicine have received the greatest financial support, which translated into a twofold increase in the number of publications. This also shows how funding influences the development of science.

In conclusion, it is worth mentioning that the motivation for writing this article was the search for academics who deal with the problem of improving decision-making processes using DS achievements. The series of these two articles that are the result of our explorations allowed us to identify the main contributors. A collaboration with the University of Leipzig has also been established, with joint research projects and scientific publications planned in the area of applying DS to business process improvement.

References:

Alejandro-Cruz, J.S., Rio-Belver, R.M., Almanza-Arjona, Y.C., Rodriguez-Andara, A. 2019. Towards a Science Map on Sustainability in Higher Education. Sustainability, 11, 3521.

Camacho, D., Panizo-Lledot, Á., Bello-Orgaz, G., Gonzalez-Pardo, A., Cambria, E. 2020. The four dimensions of social network analysis: An overview of research methods, applications, and software tools. Information Fusion, 63, 88-120.

Caviggioli, F., Ughetto, E. 2019. A bibliometric analysis of the research dealing with the impact of additive manufacturing on industry, business and society. Int. J. Prod. Econ., 208

Consoli, S., Recupero, D.R., Petković, M. 2019. Data Science for Healthcare: Methodologies and Applications. Springer.

Du, Y., Khan, H.R. 2020. Data Science for Librarians. Libraries Unlimited.

Fávero, L.P., Belfiore, P. 2019. Data Science for Business and Decision Making. Elsevier Science.

Global Data Protection Index 2020 Snapshot. https://www.delltechnologies.com/en-us/data-protection/gdpi/index.htm#.

Global Data Sphere. https://www.idc.com/getdoc.jsp?containerId=IDC_P38353.

Grus, J. 2019. Data Science from Scratch. First Principles with Python, O'Reilly.

Gryncewicz, W., Sitarska-Buba, M. 2021. Leading research by institutions and authors: A Modern Research Analysis. European Research Studies Journal, 24(3B), 1012-1026.

Lane, J. 2009. Assessing the Impact of Science Funding. Science, vol. 324.

Menzies, T., Williams, L., Zimmermann, T. 2016. Perspectives on Data Science for Software Engineering. Morgan Kaufmann.

Nymand-Andersen, P. 2021. Data Science in Economics and Finance for Decision Makers. Risk Books.

Perron, B.E., Victor, B.G., Hodge, D.R., Salas-Wright, C.P., Vaughn, M.G., Taylor, R.J. 2017. Laying the Foundations for Scientometric Research: A Data Science Approach. Research on Social Work Practice, 27(7), 802-812.

Pretty, I. 2013. Big Data in the Public Sector. Capegemini, www.in.capgemini.com/media-analysts/resources/video/big-data-in-the-public-sector.

Purnomo, A., Rosyidah, E., Firdaus, M., Asitah, N., Septianto, A. 2020. Data Science Publication: Thirty-Six Years Lesson of Scientometric Review. 2020 International Conference on Information Management and Technology (ICIMTech), 893-898.

Rao, A.S., Rao, C.R. 2020. Principles and Methods for Data Science, 43. North Holland.

Sarkar, A., Pal, A. 2019. Where does data science research stand in the 21st century: Observation from the standpoint of a scientometric analysis, 2561, 1-9.

Sarkar, A., Pal, A. 2020. Authorship Pattern of 21st Century Data Science Research: A Scientometric Evaluation, 4263, 6-15.

Thwel, T.T., Sinha, G.R. 2020. Data Deduplication Approaches. Academic Press.

Tran, B.X., Nghiem, S., Sahin, O., Vu, T.M., Ha, G.H., Vu, G.T., Pham, H.Q., Do, H.T., Latkin, C.A., Tam, W., Ho, C.S.H., Ho, R.C.M. 2019. Modeling Research Topics for Artificial Intelligence Applications in Medicine: Latent Dirichlet Allocation Application Study. J Med Internet Res., 21(11).

Van Eck, N.J., Waltman, L. 2020. VOSviewer Manual. Manual for VOSviewer Version 1.6.16. Univeristeit Leiden.