Digital Advantages for the Construction Industry

PhD candidate Plamen Yankov
University of Economics - Varna, Varna, Bulgaria
yankov.plamen@ue-varna.bg

Dr. Stefka Petrova
University of Economics - Varna, Varna, Bulgaria
s.petrova@ue-varna.bg

Chief. Assist. Prof. Dr. Svetlana Todorova
University of Economics - Varna, Varna, Bulgaria
svetlana.todorova@ue-varna.bg

Abstract

Currently, the digital transformation is recognized as a source of many potential effects for the business organization from all sectors of the economy. The purpose of this study is to identify and highlight the possible advantages of the application of digital technologies in the construction sector. A scientometric analysis of the existing literature is performed. The scope of the study is given from three perspectives through three search criteria - digitalization, big data, and forecasting. A total of 2371 articles are abstracted. Then, the extracted data is visualized through Vosviewer software tool. The growth of publications increases significantly over the last decade. The results illustrate that the digitalization in the construction sector affect all aspects of construction projects, with the strongest impact on the architecture design and building information modelling. Big data in construction is associated with the data storage, data analytics and information management, during the whole life of the buildings. The third search criterion shows that construction companies most often forecast the total costs using regression analysis, machine learning algorithms, artificial neural networks, etc The research findings could support decision makers and practitioners with-depth understanding for the possible advantages of digital technologies in the construction industry. The current study is part of a larger project called "Digitalization of Economy in a Big Data Environment" BG05M2OP001-1.002-0002-C02.

Keywords: construction industry, big data, digitalization, forecast, scientometric analysis, Vosviewer

JEL Code: E020

Introduction

The development and use of ICT both lead to the growth of vast amounts of data which are better known as big data. The term "big data" is first used in October 1997 by two NASA researchers, Michael Cox and David Ellsworth. In their article, they name the emerging challenges facing computer systems, mostly caused by voluminous video files and other datasets, which are usually so large that they hamper the capacity of main memory, local disks, and even remote server space. They call this the big data problem (Cox and Ellsworth, 1997).

The consulting company Gartner characterized big data by three main characteristics: volume, speed and variety, such as the so-called 3 cV's model (figure 1).

At present, researchers from all over the world are adding new features to big data and already known over 51 V's characteristics in addition to the main 3 (Khan et al., 2019). Each of the characteristics describe a specific component of the big data, for example variability is recognized as dynamic, evolving behaviour in sources. The first 30 of these new features are listed in Table 1.

Over the last few years, scientists, academic researchers and business practices are focused on the possibilities for identifying useful insights and knowledge from big data (Petrivskyi et al., 2020). The digital transformation (Petrov&Nacheva, 2020) in the construction industry expresses rather weakly in comparison with the IT, financial, commerce, tourism, logistics and some other economic sectors (Doroshenko, 2020). At the same time, many business organizations in the construction industry apply new technologies (Panayotova et al., 2016) in order to gain advantages

over the competitors. Building Information Modelling (BIM), Internet of Things (IoT), cloud computing are widely used in the sector (Garyaev and Garyaeva, 2019). The main benefit of digitalization in the construction industry is the increase of productivity (Calvetti, 2020). These include increasing productivity and quality, as well as reducing overall costs, delays and accident risks.

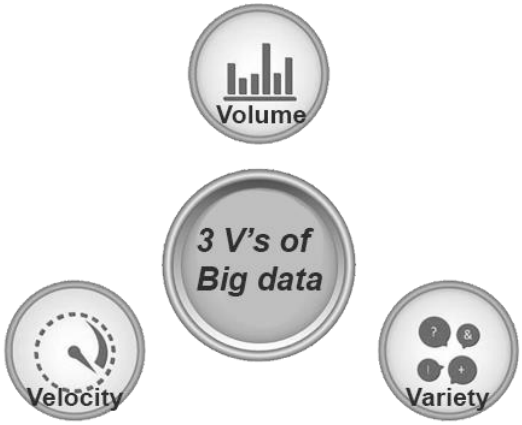

Figure 1. Main characteristics of big data; source – adapted from Internet.

Table 1. Additional 30 characteristics to big data.

source - adapted from (Khan et al., 2019)

№	Big data characteristic	№	Big data characteristic	№	Big data characteristic
1	VOLUME	11	VERBOSITY	21	VAGUENESS
2	VELOCITY	12	VERSATILITY	22	VITALITY
3	VARIETY	13	VOLUNTARINESS	23	VIRALITY
4	VERACITY	14	VIRTUALIZATION	24	VALOR
5	VALUE	15	VARIABILITY	25	VANE
6	VALIDITY	16	VISCOSITY	26	VANILLA
7	VOLATILITY	17	VOCABULARY	27	VANTAGE
8	VIABILITY	18	VENUE	28	VARIFOCAL
9	VISUALIZATION	19	VIOLATION	29	VARMINT
10	VERIFICATION	20	VERSIONING	30	VARNISH

Building Information Modeling (BIM) is at the heart of the construction industry's digital strategy and also has the ability to create vast amounts of digital data (Hyder and Bandi, 2021)(Yin *et al.*, 2019)(Dimitrov et al., 2016) (Kostadinova et al., 2016). For example, a three-story building model can exceed 50 GB in data volume (Bilal *et al.*, 2016). For datasets larger than 5TB, companies typically use new, customized software. Well known representatives of this type of software products are: Map Reduce, Big Table and Apache Hadoop. At the same time, Hadoop can also be used to store BIM files. In this regard big data is useful for the construction industry as a platform to derive potential insights from the accumulation of BIM digital data (Hyder and Bandi, 2021). The concept of connected devices that exchange data in real time is known as the Internet of Things (IoT) (Agarwal and Dhar, 2014). Iot could be also described as s system that incorporates various independent systems into a single whole (Adamu *et al.*, 2020). In the context of smart homes, many of the installations such as electrical, cooling, heating and security are built at the construction level (Alaloul et al., 2019). This trend is also a big data source for the construction companies. The interconnection of individual machines, smart devices and sensors can also

facilitate the control of construction projects, which is often a challenge (Alexandrova, 2020). Miglena Stoyanova emphasizes the growing importance of big data for property management ensuring a more efficient experience for professionals and customers (Stoyanova, Vasilev and Cristescu, 2021). The growth in the number of software applications used in construction and their improvement in general leads to an increase in the volume of data (Alexandrova, 2020). Snezhana Solova explores the need for new innovative data management models in the construction sector (Sulova, 2020). The benefits of cloud technology are widespread in the construction industry, where large BIM files often have to be accessed remotely (Sulova, 2020).

1. Methodology

The Scientometric analysis conducted in this study is organized in three search directions where each of them follows the algorithm from table 1 bellow.

Table 2 Scientometric algorithm; source – the author

№	Step	Description
1.	Search criteria selection	Formulation of search criteria and analysis of data by sources and bt years
2.	Data collection	Collect bibliographic data from Scopus and download it into a .csv file
3.	Visualize	Create a bibliometric co-occurance keywords network with Vosviewer software tool
4.	Pattern recognition	Extract keywords and sort them by total link strength

The Scopus database is chosen for this study as a source of publications on the research subject. It is one of the largest and most reputable scientific databases. Furthermore, only journal articles are selected for analysis, as journal articles usually provide more comprehensive and high-quality information than other types of publications. As an additional advantage Scopus also has a built-in useful feature which allows easily download bibliographic data into a .csv file.

Vosviewer is a free, qualitative bibliometric software tool which is user friendly and easy to navigate. It automatically assigns keywords into clusters based on their co-occurrences in the extracted bibliographic data. In Vosviewer, the size of an element depends on its degree of repeatability in the extracted publications, the strength of the connection, and so on (Wu et al., 2019). The individual elements with an available connection between them form one cluster, and the different clusters are presented in different colours. After each co-occurring network the first 20 key words with the higher total link strength are listed in a table for better differentiation. According to the Vosviewer manual, each link has a strength, represented by a positive numerical value. The possible areas of application and potential advantages are analysed from these three perspectives together.

2. Digitalization in construction industry

The first round of paper retrieval is conducted with the following search criteria: TITLE-ABS-KEY= (**digital*** AND **"construction industry"**). The time range is set to all years which includes publications from 1960 to present. Digitization is a process that begin with the introduction of the first computer configurations. The search results are presented in Figure 2.

Interestingly, the first document "Earthwork and rock analysis computer program for building sites" date back to 1973. Unfortunately, this paper as well a few other elder publications are not accessible. Approximately 10 years later in 1984 there are 8 publications. Some of the titles of these scientific papers are: "Use of simulation models in construction", "Exploring the potential for automated real-time data acquisition in construction", "Digital training simulators in construction", etc. The growth in publications on the topic is illustrated in figure 3 below.

Figure 2. Search criteria and results in Scopus.

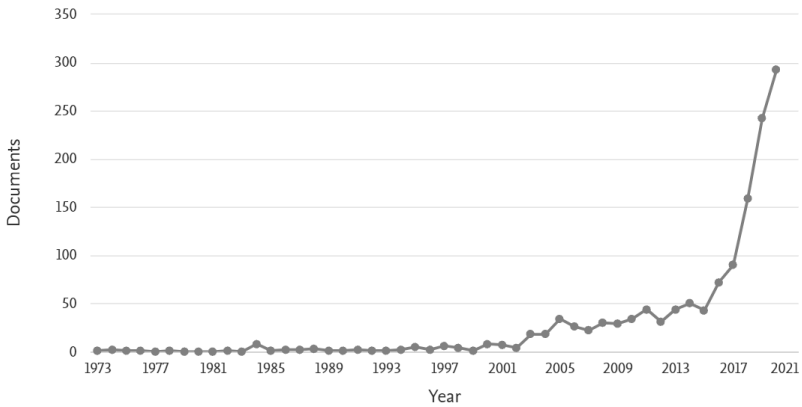

Figure 3. Publications by year with search criteria (digital* AND "construction industry") by year; source - Scopus;

Most of the retrieved publications belong to the field of engineering (38.4%), followed by computer science (18.4%) and business, management and accounting (7.2%). This proves that the analysed literature is relevant to the purpose of this study. The sources of the analysed publications are shown in figure 4.

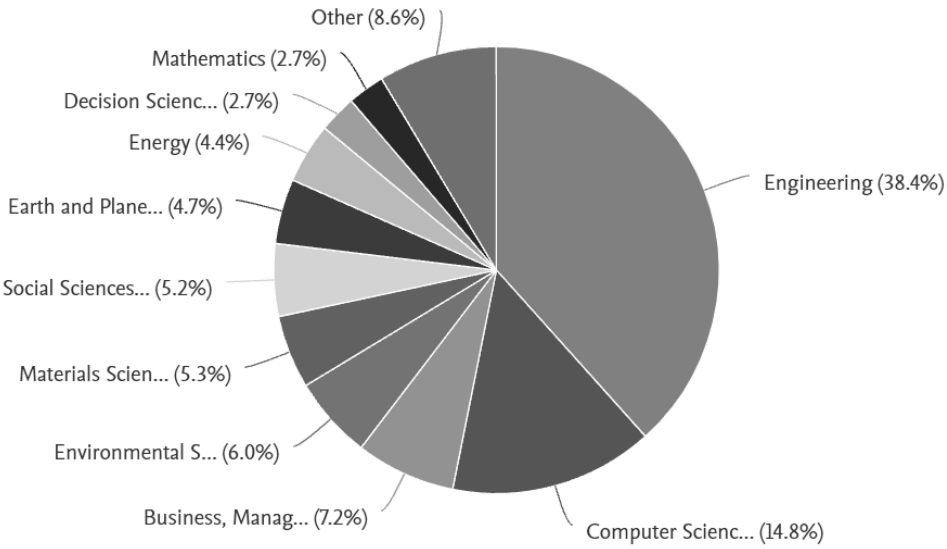

Figure 4. Documents by source; source - Scopus.

The extracted data from Scopus contains information about title, abstract, authors and keywords. In the specific case of this study, the information is visualized according to the keyword criterion. The most frequently cited keywords and the links between are illustrated in figure 5.

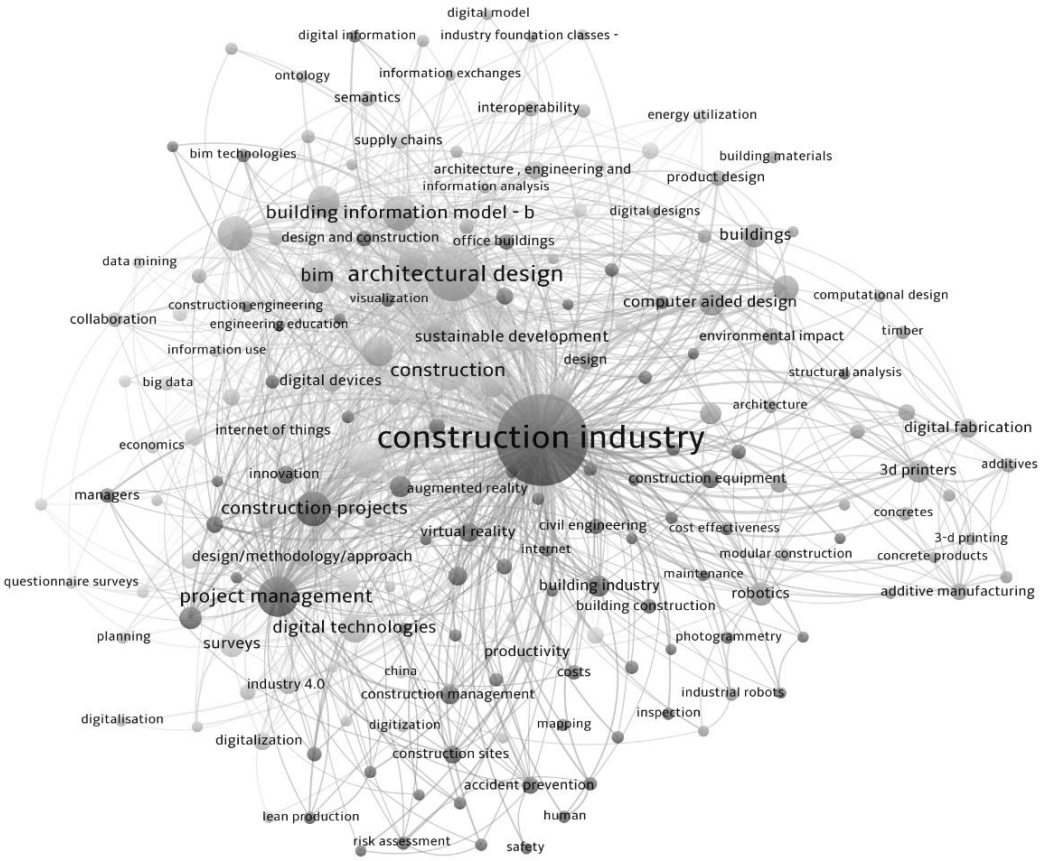

Figure 5. Network of co-occurring keywords; bibliographic data with search criteria (digital* AND "construction industry")

The details of the first 20 most recurring keywords in total link strength are presented in table 3 below. The total link strength indicates the number of publications in which two keywords occur together (Van Eck and Waltman, 2018). It is interesting to find from the table that digitalization has the strongest influence in the building information modelling (BIM), architectural and structural design and construction projects in general.

Table 3 Details for the total link strength of the top 20 occurrence keywords

№	keyword	total link strength
1	construction industry	6537
2	architectural design	3223
3	construction	1685
4	project management	1547
5	building information model - bim	1344
6	information theory	1239
7	building information modelling	1196

8	construction projects	1123
9	digital storage	1065
10	information management	992
11	life cycle	944
12	bim	915
13	digital technologies	673
14	structural design	649
15	sustainable development	627
16	building information modelling	576
17	computer aided design	567
18	surveys	544
19	robotics	540
20	construction sectors	503

3. Big data

The second paper retrieval is conducted with the following search criteria: TITLE-ABS-KEY = (**"big data"** AND **"construction industry"**) In this round a total of 195 articles are retrieved as it is shown in figure 6.

Figure 6. Search criteria and results in Scopus.

For this search, both arguments are in quotation marks to narrow the study and achieve higher accuracy. Otherwise, this search would generate results that relate to big data in construction and other industries unrelated to the subject of the current study. The dynamics in publications of the second search (figure 7) confirms that big data application in the construction industry is an emerging topic which grows exponentially after 2014.

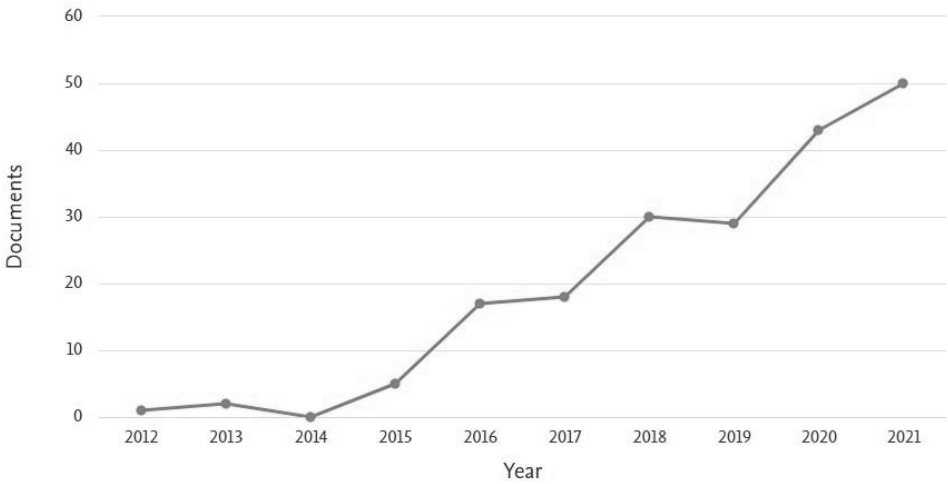

Figure 7. Publications by year : "big data"AND "construction industry", total 195 articles; source – Scopus;

The most publications belong to the engineering (28.7%), followed by computer science (16.8%), business, management, and accounting (8.4%). The full distribution by source is presented in figure 8.

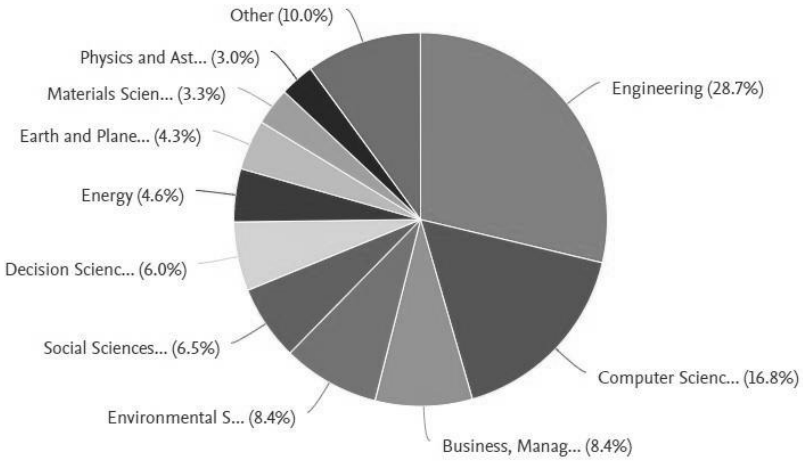

Figure 8. Publications by source.

For the next network visualization, the minimum keyword citations are set to 3. From figure 7 it can be seen that that all keywords are divided in 5 clusters according to color.

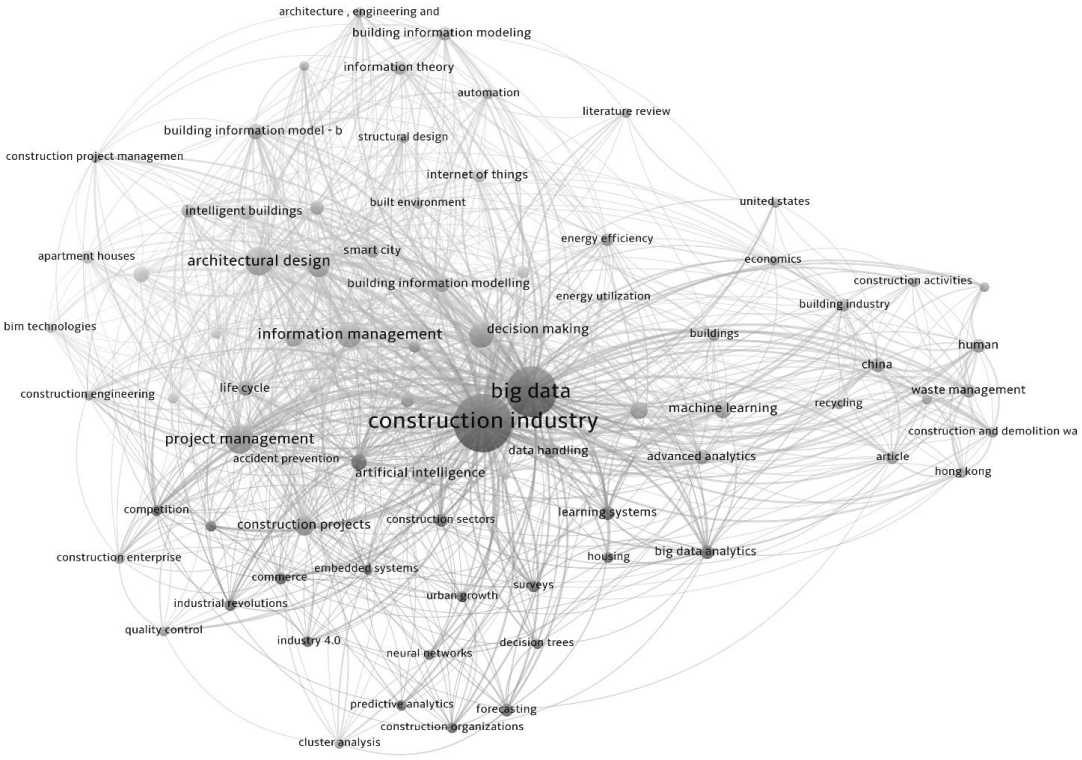

Figure 9. Network of co-occurring keywords.

The first 20 keywords with highest total link strength are presented in table 4. As implied by the search criterion, logically the table contains many data sources such as BIM, architectural plans, life cycle, but at the same time there are many methods and techniques for data collection and analysis.

Table 4 The total link strength of the top 20 occurrence keywords

№	keyword	total link strength
1.	construction industry	826
2.	big data	671
3.	architectural design	280
4.	project management	257
5.	information management	207
6.	data mining	190
7.	construction	164
8.	construction projects	123
9.	data analytics	113
10.	artificial intelligence	108
11.	intelligent buildings	104
12.	building information model - bim	103
13.	digital storage	102
14.	decision making	101
15.	machine learning	95
16.	construction companies	92
17.	building information modelling	90
18.	life cycle	89
19.	BIM	87
20.	China	81

From table 4 is noticeable that the most frequently mentioned keywords are either directly related to the data - such as: big data, information management, data mining, data handling, digital storage, cloud computing or are related to data techniques and methods for data management such as: artificial intelligence, machine learning, decision making, classification, learning systems, Based on general similarities, the first 20 most commonly used keywords can be classified into two areas of application, as grouped in Figure 10.

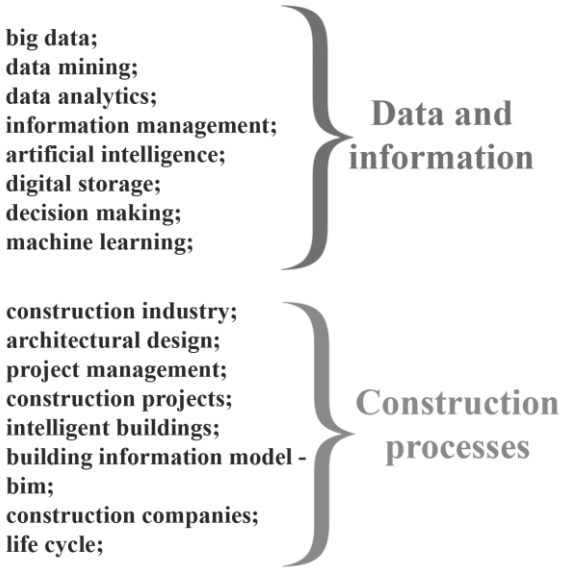

Figure 10. Classification of the 20 most frequently co-occurrenced keywords

As mentioned in the introduction, BIM technology is one of the big data generators for construction and for this reason the term is present in parallel with big data. Big data play a vital role in forecasting in the sector (Ngo, Hwang and Zhang, 2020). Big data applications are typically associated with large numbers of users, different software systems, and storage architectures (Shahriari *et al.*, 2016).

4. Forecasting

The third direction of research in this paper is the application of methods for forecast and predictive analysis in the construction industry. The growing importance of the digitalization and big data for the construction industry is often recognized with the possibility of identifying critical insights and more accurate forecasts. The third search is executed with the following criteria: TITLE-ABS-KEY = (**predict*** AND **forecast*** AND **"construction industry"**). In the scientific literature, the words forecasting and predicting are often used to calculate future results. For this reason, both are included as search arguments. With this search, a total of 535 bibliographic records are collected – figure 11.

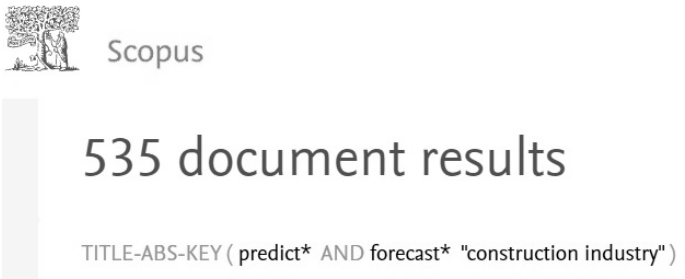

Figure 11. Search criteria and results in Scopus.

The first publications in Scopus dates from 1978. In the last few years this area also notes an high interest due to the increased amount of data and the possibility for more accurate forecasts

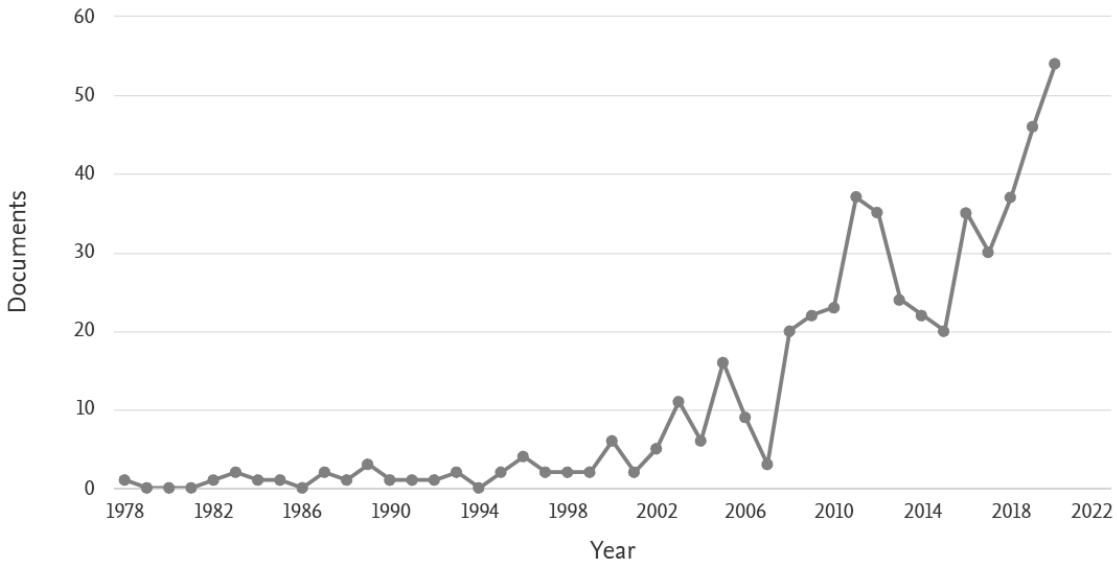

Figure 12. Publications by years.

From figure 13 is noticeable that again the most publications are in the field of engineering. The second group of sources are from business, management and accounting due to the need for models and algorithms for forecasting.

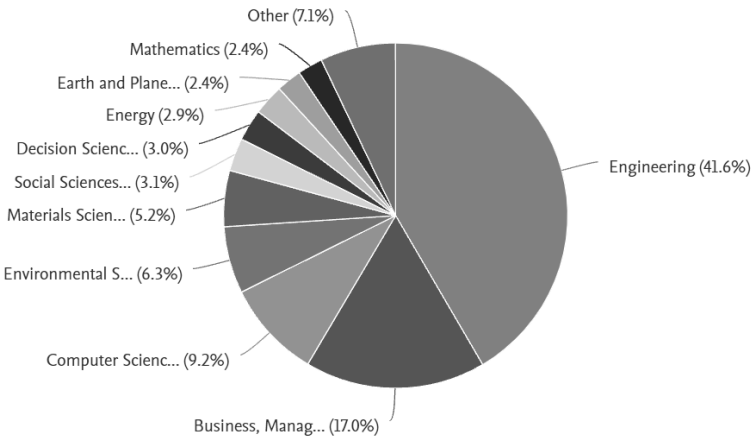

Figure 13. Publications by source.

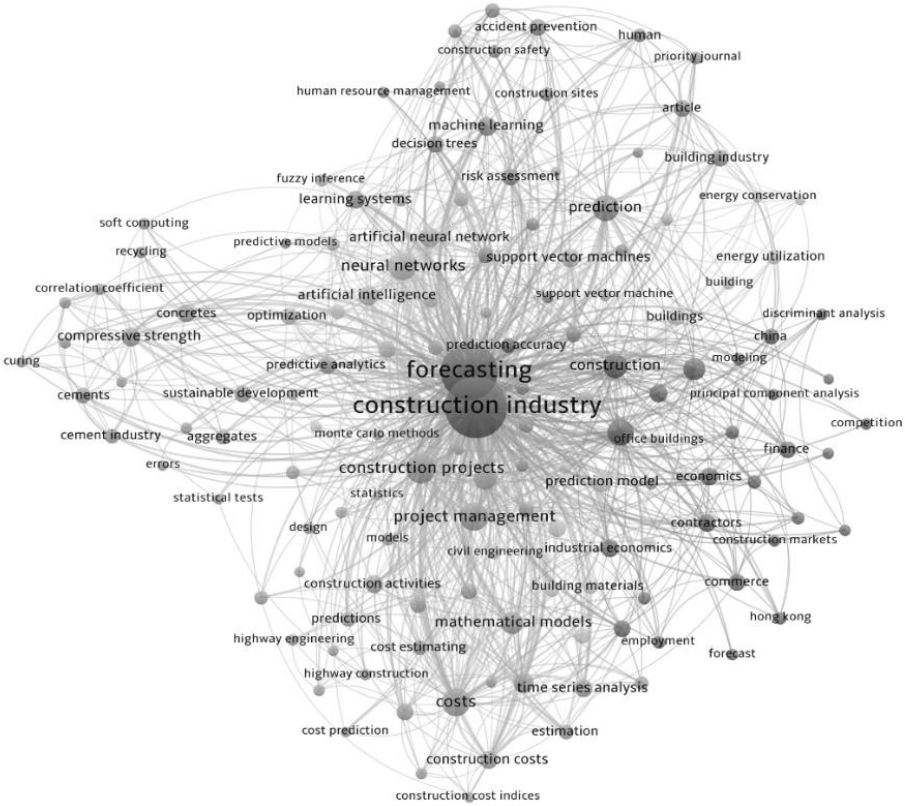

Figure 14. Network of co-occurring keywords.

The details of the keywords recurring are shown in Table 5. The number of keywords corresponds to the top 20 most recurring ones in the network in Figure 14.

Table 5. The total link strength of the top 20 occurrence keywords.

№	keyword	total link strength	№	keyword	total link strength
1	construction industry	2917	11	mathematical models	307
2	forecasting	2548	12	cost benefit analysis	305

3	construction projects	654	13	artificial intelligence	293
4	neural networks	608	14	construction companies	274
5	costs	573	15	artificial neural network	269
6	project management	568	16	support vector machines	254
7	regression analysis	512	17	prediction model	253
8	prediction	462	18	machine learning	251
9	construction	394	19	construction costs	244
10	decision making	323	20	compressive strength	219

Among the main application of big data in the construction industry are decision making, forecasting, cost estimation, architecture design, Linear regression models are the most basic machine learning algorithms (Sarker, 2019). Some of the construction processes are seasonal and for this reason forecasting with autoregressive models with integrated moving average (Kim, Abediniangerabi and Shahandashti, 2020). Researchers emphasize the importance of models for calculating total construction costs as a preventive measure to avoid cost overrun (Mohammad, Mostafa and Nassar, 2018) (Gotlur et al., 2020) (Plebankiewicz and Wieczorek, 2020). Among the most common models for predictive analysis are decision trees, regressions (linear and logistic) and neural networks (Halton, 2019). Neural networks are also used for cluster data and information classification.

Conclusion

In the last few years, the construction industry successfully implements new digital technologies. This study investigates the capabilities of the digital transformation for the construction industry. Scientometric approach is applied, and analysis is performed in 4 steps. They include search criteria formulization, bibliographic data extraction, co-occurrences keywords visualising and pattern recognition. The bibliographic data are extracted from Scopus. The majority of the analysed literature is in the field of engineering, computer science and business and management. Vosviewer software tool is used for network reoccurring keywords mapping. Based on the visualized clusters of keywords and tables with the most frequently repeated keywords, several important conclusions are made. Digital technologies applied by construction companies are focused mainly on the architectural design of buildings and construction projects in general, information management and achieving sustainable development. Modern design software solutions, BIMs and various sensors in smart homes accumulate big data that are used from the construction companies for decision making. Big data arrays often require processing with machine learning and deep learning techniques. Last but not least, this study emphasizes that the main subject of forecasting in construction is the cost of projects and the most used methods for their calculation are regression analysis, various machine learning techniques and artificial neural networks.

References

1. Adamu, A. A. *et al.* (2020) ‘An integrated IoT system pathway for smart cities’, *International Journal on Emerging Technologies*, 11(1), pp. 1–9.
2. Agarwal, R. and Dhar, V. (2014) ‘Big data, data science, and analytics: The opportunity and challenge for IS research’, *Information Systems Research*, 25(3), pp. 443–448. doi: 10.1287/isre.2014.0546.
3. Alaloul, W. S. *et al.* (2019) ‘Industrial Revolution 4.0 in the construction industry: Challenges and opportunities for stakeholders’, *Ain Shams Engineering Journal*. THE AUTHORS, (xxxx). doi: 10.1016/j.asej.2019.08.010.
4. Alexandrova, Y. (2020) ‘ИНОВАТИВНА ЦИФРОВА ОБРАБОТКА НА НОВИ И СЪЩЕСТВУВАЩИ ДАННИ INNOVATIVE DIGITAL PROCESSING OF NEW AND

EXISTING DATA', pp. 92–120.
5. Bilal, M. *et al.* (2016) 'Big Data in the construction industry: A review of present status, opportunities, and future trends', *Advanced Engineering Informatics*. Elsevier Ltd, 30(3), pp. 500–521. doi: 10.1016/j.aei.2016.07.001.
6. Calvetti, D. (2020) 'Worker 4 . 0 : The Future of Sensored Construction Sites', pp. 1–22.
7. Cox, M. and Ellsworth, D. (1997) 'Application-controlled demand paging for out-of-core visualization', *Proceedings of the IEEE Visualization Conference*, (July), pp. 235–244. doi: 10.1109/visual.1997.663888.
8. Dimitrov, G., Panayotova, G., Garvanov, I. et al. (2016). Performance analysis of the method for social search of information in university information systems.. 3rd International Conference on Artificial Intelligence and Pattern Recognition (AIPR), Lodz, Poland, IEEE, 2016, pp.149-153.
9. Doroshenko, A. (2020) 'Applying Artificial Neural Networks in Construction', *E3S Web of Conferences*, 143, pp. 2018–2021. doi: 10.1051/e3sconf/202014301029.
10. Van Eck, N. J. and Waltman, L. (2018) 'VOSviewer Manual: Manual for VOSviewer version 1.6.7', *Univeristeit Leiden*, (February), p. 51. Available at: https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.8.pdf.
11. Garyaev, N. and Garyaeva, V. (2019) 'Big data technology in construction', *E3S Web of Conferences*, 97. doi: 10.1051/e3sconf/20199701032.
12. Hyder, F. and Bandi, S. (2021) 'Associations Between Building Information Modelling (BIM) Data and Big Data Attributes', pp. 1–11.
13. Khan, N. *et al.* (2019) 'The 51 V's of big data: Survey, technologies, characteristics, opportunities, issues and challenges', *ACM International Conference Proceeding Series*, Part F1481, pp. 19–24. doi: 10.1145/3312614.3312623.
14. Kostadinova, I., Toshev, R., et al. (2016). Temporal Analysis of the Pedagogical Adoptions use and Application of the Augmented and Virtual Reality Technologies in Technical Subject Areas. 11th Annual International Conference of Education, Research and Innovation, ICERI2018 Proceedings, Seville, Spain : IATED, 2018, pp.4387-4393.
15. Ngo, J., Hwang, B.-G. and Zhang, C. (2020) 'Factor-based big data and predictive analytics capability assessment tool for the construction industry'.
16. Panayotova, G., Dimitrov, G., et al. (2016). Modeling and data processing of information systems. 3rd International Conference on Artificial Intelligence and Pattern Recognition (AIPR), Lodz, Poland, IEEE, pp.154-158.
17. Petrivskyi, V., Dimitrov, G., Shevchenko, V., et al. (2020). Information Technology for Big Data Sensor Networks Stability Estimation. Information and Security, Sofia: Procon Ltd. 47(1), pp.141-154. https://doi.org/10.11610/isij.4710
18. Petrov, P., Nacheva, R. (2020). Информационни системи за социална бизнес аналитичност в реално време. Варна: Наука и икономика. Библ. Проф. Цани Калянджиев, Кн. 61.
19. Shahriari, B. *et al.* (2016) 'Taking the human out of the loop: A review of Bayesian optimization', *Proceedings of the IEEE*. IEEE, 104(1), pp. 148–175.
20. Stoyanova, M., Vasilev, J. and Cristescu, M. P. (2021) 'Big data in property management'.
21. Sulova, S. (2020) 'Юбилейна международна научна конференция ИКОНОМИЧЕСКА НАУКА, ОБРАЗОВАНИЕ И РЕАЛНА ИКОНОМИКА: РАЗВИТИЕ И ВЗАИМОДЕЙСТВИЯ В ДИГИТАЛНАТА ЕПОХА Jubilee International Scientific Conference ECONOMIC SCIENCE, EDUCATION AND THE REAL ECONOMY: DEVELOPMENT AND INTERA', (August).
22. Yin, X. *et al.* (2019) 'Building information modelling for off-site construction: Review and future directions', *Automation in Construction*, 101(January), pp. 72–91. doi: 10.1016/j.autcon.2019.01.010.