

Big data as a business opportunity: an Educational Perspective

Ilona Pawełszek
Częstochowa University
of Technology,
Management Faculty
Poland
Email:
ipaweloszek@zim.pcz.pl

Jędrzej Wieczorkowski
Warsaw School of Economics
Institute of Information Systems
and Digital Economy
Poland
Email:
jedrzej.wieczorkowski@sgh.waw.pl

□ **Abstract**—The paper presents considerations on the concept of big data. The aim of the paper is to confront the attempts of defining big data with its common understanding by different groups of users. In this research the group of respondents are students of Warsaw University of Economics. The authors advocate that the student's opinion and attitude to big data can be important regarding the fact that they will probably become managers - business users of big data solutions. Another aim is the analysis of present educational offerings of Polish universities in the area of big data and suggesting the directions of the development of educational curricula to better fit the marked demand for big data professionals.

I. INTRODUCTION

THE attempt to demystify Big Data may on one hand lead to numerous definitions that can be found in academic literature, corporate and individual blogs of companies and people interested in latest advances, algorithms, and practical applications in this field [1] [2]. On the other hand some publications regarding better common understanding of Big Data and its social aspects can be found [3].

The term 'big data' is very simple in its construct, it is a combination of two frequently used words 'big' and 'data'. The simplicity of the term makes even non-specialists have

their own idea on what big data really is. The term suggests gathering, processing and using large volumes of data.

'Big' is a relative term which changes its meaning along with technology development. Particularly performance and capacity challenges that are associated with big data systems evolve over time in their nature, scale and scope. Thus it is not surprising there is much confusion surrounding the term.

The phrase 'big data' has been extensively used in the recent years. For example analysis of keyword popularity with the Google Trends tool reveals the constant, systematic and regular growth of number of searches with the keyword 'big data'. This phenomenon has been taking place since the year 2012 (see. Fig.1). However the question about context of the searches stays open so the further research is sorely needed.

Over the past five years, the emergence of huge data sets and data-intensive science are fundamentally changing the way researchers work in virtually every scientific discipline. This phenomenon can be easily observed while analyzing bases of scientific publications (see. Fig.2.).

Big Data is currently understood as the term increasingly used to describe the process of applying serious computing power to seriously massive and often highly complex sets of information [4], data sets whose sizes exceed the capacity of

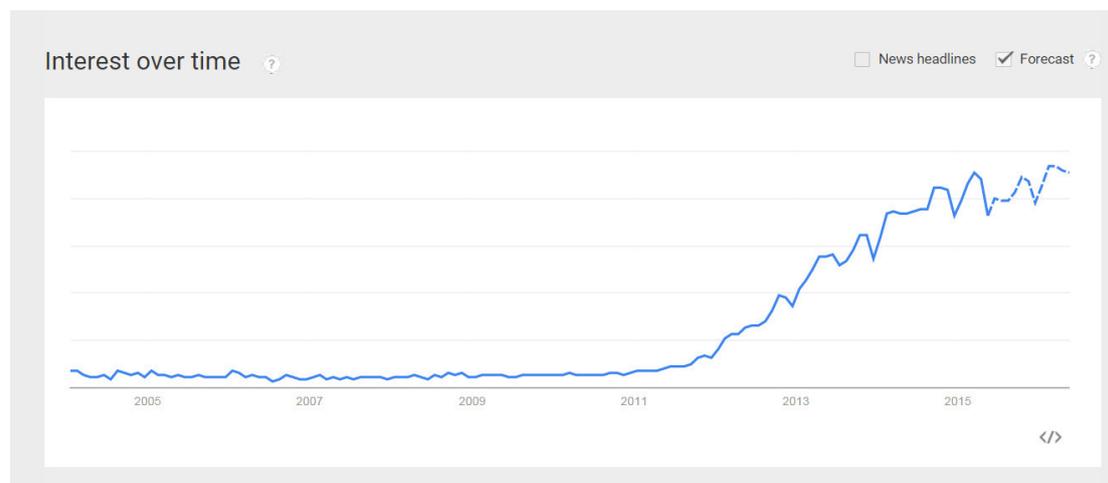


Fig. 1 Big data search term – interest over time with forecast, Source: Own elaboration

conventional database tools for gathering, storing, managing and analyzing data [5]. The above definitions are timeless and universal because they are not domain-specific and do not specify the typical big data features in any point of time, however they emphasize the change that has happened as compared to ‘conventional database’. This point of view is highlighted for example by Dumbill [6] who states that: Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn’t fit the strictures of database architectures.

Other examples of definitions underline the unstructured character of data, i.e. accordingly to Rouse [7] big data is a general term used to describe the voluminous amount of unstructured and semi-structured data a company creates - data that would take too much time and cost too much money to load into a relational database for analysis.

Model 3V of META Group [8] is usually considered a starting point for describing big data in the modern sense. The big data notion was not considered then, but its 3 basic features were taken into account: Volume (very large volumes of processed data), Velocity (variability and dynamics of processed data), Variety (variety of processed data). Later various authors tried to choose other features characteristic for big data, which could enhance the ‘v’ model, particularly: Value – the processed data is valuable, Veracity – another way ‘accuracy’ or ‘truthfulness’, Visualisation – the possibility to present the vast amount of data in a way that would be comprehensible, easy to understand and read.

The ‘Volume’ feature is the most difficult to verify among the original 3V because the amount of data significantly increases with the IT development. The volume of large datasets may differ depending on the area of application. The remaining ‘v’ features seem very important: the possibility of real-time analyses of particular variables, such as stream data, sensor data (velocity) and analysis of various unstructured data (variety). These features are characteristic to contemporary big data in contrast to the previous approaches to massive data processing.

Attempts to define the term big data are as well very general and timeless as bringing up the technological aspects and trends which on one hand describe the concept in a more detailed way but on the other hand go obsolete very quickly. At the same time, because of dynamics of the discussed term different opinions can be found agreed that the concept is not yet fully definable.

II. THREE-ASPECT APPROACH TO BIG DATA. THE AUTHOR’S WORK ON UNDERSTANDING OF BIG DATA CONCEPT

The lack of unified definition of the term big data coupled with the high recognition and popularity as well in scientific papers as in popular publications induced the Authors to investigate its common understanding by different professional and social groups.

The Authors conducted research on the context of using the term ‘big data’ in press articles from daily newspapers [3]. The analysis of online edition the newspaper of Agora (one of the biggest media companies in Poland) was conducted twice: in March and September 2014. The thematic context of 10 of the top search results for the key word ‘big data’ had been analyzed after excluding the incidental cases where the document’s context was not relevant to the research subject.

The Authors proposed “three-aspect approach”, identifying three essential aspects of big data: technological, business, social.

The technological aspect in the above classification represents a focus on the methods of big data analysis and information technology used. Considering the methods of analysis the big data concept is based first of all on well-grounded statistical methods, artificial intelligence, machine learning and data mining. Similarly to the latter, the aim of big data analysis can be exploration of implicit associations and patterns in large data sources. The big data approach is to analyze all the accessible transaction data rather than a representative sample. Such approach does not require a predefined model and hypothesis, but enables discovering

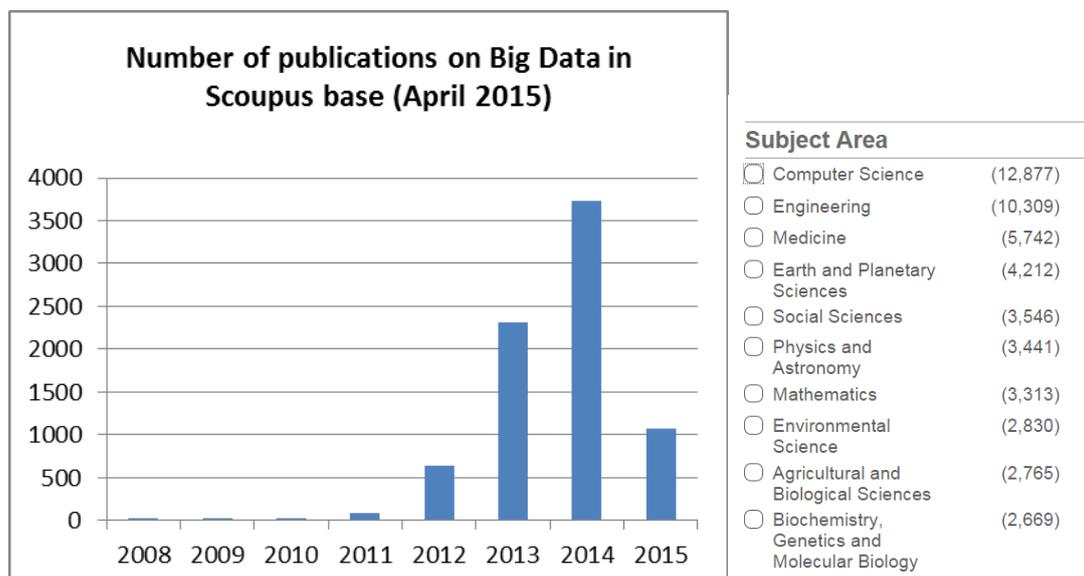


Fig. 2 Number and subject areas of scientific publications on big data in Scopus base, Source: Own elaboration

previously unknown patterns in all the accessible data. Moreover the big data approach gives a possibility to analyze unstructured data such as texts from internet pages.

The Business aspect of the described classification focuses on applications of big data. Considerably shrinking cost of data gathering and processing makes it worth to process the data which were previously not possible or economically not feasible. The new approach can be the solution for information overload which refers to the situation when there is enough accessible and relevant data but there are no capabilities to process or draw conclusions on the basis of the data [9]. The big data can be considered the next stage in evolution of business intelligence and predictive analysis. In this context the possibility of using transactional data directly for analysis seems to be very important qualitative change. This change is especially viable for implementation of new organizational solutions and methods. As Płoszajski [10] notices in history it was always so that when the prevalence replaced rarity, the necessity for new business models arose.

Part of the new applications are old but improved solutions for data mining, risk modelling, fraud detections, churn predictive modelling, load forecasting for electric power systems, stock technical analysis, optimization of logistics [11]. Combining the concept of big data with other ideas (e.g. GIS, cloud computing) allows many new benefits, for example in supporting public administration, regional electronic communities [12].

The social aspect of big data is in the Authors' opinion associated with social consequences of data processing results. These matters are often overlooked in definitions of big data concept. Alongside the clear and unquestionable advantages for society, there are also some threats, the first of which is associated with massive processing and using of personal data. In contemporary world every person leaves large amount of electronic traces such as records in the data bases or different archives of unstructured data. These are the data describing the activities of the internet user, in particular e-mails, posts on social networking websites, pages visited, keywords and phrases input into search engines, billings and geolocation data from mobile networks, data on financial transactions (payments with credit cards, bank transfers), data from loyalty programs, purchases, medical sensitive data, video from monitoring systems. The abovementioned data may have great commercial value (for example enable personalized targeting and displaying custom ads) and public value (i.e. protection of public security). The essential issue in question is the required level

of privacy which can be different in commercial and public applications of big data processing.

The Author's researches on the concept of big data were focused on assigning the context to the press articles and notes. Because some of the papers were focused on several aspects, the total number of references in each text was over 10. The summarization of research results is presented in table I.

It should be highlighted that the research encompassed only nonprofessional press. Moreover what is puzzling is a very small number of references to technological aspect. The articles which were strictly economical were focused on the examples of big data applications. Most of the references were to the social aspect. The potential threats of privacy were often discussed especially in the contest of data processing by various institutions, the control of the internet and social networking websites by commercial units.

It is generally can be seen in the press articles that particularly interested for journalists is the social aspect, to a lower degree applications and least interesting were the technological and analytical issues. The reason for this lack of interest in technological aspects is probably the lack of expertise and competencies in technology and data processing. Most of the journalists were trying to follow popular, trendy topics. Popularizing big data the journalists often emphasize the influence of technological developments and economic changes on society.

III. RESEARCH ON KNOWLEDGE BIG DATA CONCEPTS

The Authors decided to explore the knowledge and understanding of big data among students. The aim of the research was also to verify the hypothesis that the social aspect of new technologies is particularly important for the respondents. The studied community was a group of students from the Warsaw School of Economics.

Apart from openness to the new solutions, which is the typical feature of young people, it is vital to be broadminded and realize the potential of the new technologies applications. In particular the students of economic schools should have well-grounded, good idea about ICT applications in widely understood economy. This issue is especially important for this research. The future users of big data – based solutions will probably be today' students of economics, statisticians and analysts (so called data scientists). A broad group of big data users will probably be the management professionals from companies and institutions who will commission studies of big volumes of data. Therefore the group of respondents was intentionally chosen to draw conclusions about the desirable new directions for higher education in the area under consideration.

The group of 140 students underwent the survey using traditional paper questionnaires with closed questions. Among the surveyed persons 75 were first-year students and 65 second of third-year students. In the Warsaw School of

Table I. Research on the concept of big in the press articles and notes

Period	Aspects		
	technological	business	social
March 2014	1	5	8
September 2014	3	5	7

Source: Own elaboration

Economics the first-year students implement the standard study program including, inter alia, basic courses on technology and statistics.

In the next years of studies the students have more possibilities to choose the courses and their sequence depending on their interests. For this reason in the research procedure students are divided on two groups – the first year of study and higher (II and III) years of studies. The respondents were asked to state whether they are acquainted with the term ‘big data’ and more detailed questions regarding the understanding or the considered term (11 questions) and its characteristic features (8 questions) were asked. The data were gathered during the educational activities not relevant to big data concept. Such a form of survey ensured virtually 100% response rate.

The assumption was made that the students have enough knowledge on the basic level to understand the capabilities of modern information technologies and their applications (in particular in business), and also they should realize the social consequences of using big data.

For the question „Have you ever met the term ‘big data’ in contest of new possibilities of massive data processing?” there were 58% of positive answers. It should be noted that the number of positive answers is increasing. Among the students of the first year there were 40% of positive answers, while on the higher years of studies there was 77%.

Although it cannot be claimed clearly that the subject of big data is included in the obligatory courses for bachelor studies, the teachers may mention about it during the obligatory course of ‘Statistics’ on the first year of study. Moreover the concept can appear in the content of other courses however it is not fully discussed.

In this case the understanding of big data term is of special interest. The further analysis of survey results will only take into account the students who have acquaintance with the term big data. A clear majority of respondents (68%) understands big data concept very broadly, as every processing of large amounts of data. This group does not think of big data in terms of theoretical considerations and it can be supposed that they understand the term ‘big’ in a purely common way. However regarding the continuous theoretical considerations on the term big data and the lack of uniform definition it cannot be assumed that this ‘shallow’ understanding is not correct.

The students’ faith in the efficiency of typical big data methods can be observed. Most of them (65%) claim that the characteristic feature of big data is quality of data processing: 67% of the first year students, and 62% of the higher years of study. In consequence most of the students do not see the problems of data quality which are in fact typical to big data. And again the awareness of importance of this problem increases along with the year of studies (17% totally, in which 7% for the first year and 24% for the second and third year).

The majority of the respondents stated that great business value of processed data is characteristic for big data (83% of all the respondents, 77% first year of study, 86% higher year of study). Because the respondents are students of economics it seems natural that they notice the significance of detailed data in management. However it can be noted that the students’ attitude to the massive data processing tends to be slightly indiscriminate. And again the criticism increases long with the education level and is probably caused by improving knowledge. Considering the field of study, it can be assumed that the business aspect of big data is more clear for the respondents than the technological one. The technological issues are not explained in detail, similarly the analytical methods are discussed only on the selected study specializations on master degree.

The survey results on social aspect of big data are also interesting. According to most of the students a threat to privacy when using massive data processing for surveillance, monitoring of individuals and identification are not associated with big data (45% of positive answers). This is also the case with privacy law and personal data processing with IT tools (35%). Generally the number of positive answers is not low. Compared with the previous research on journalists, the students pay more attention to possibilities of using big data for economic purposes than on threats to privacy. It can be assumed that young people somehow used to partial loss of privacy, due to the fact they extensively use internet, in particular social networks that is necessarily associated with privacy issues. However to prove the hypothesis it would be necessary to conduct specific studies.

The above discussed results of the survey are related primarily to questions about the opinion: how wide is the concept of big data, which is practical importance and what could be the negative consequences of applications. Other questions allow for verification of the correct understanding of the concept. Most respondents correctly classified technology of high-performance computers into big data idea (62%) and very large amount of processed data (91%). These questions do not relate to technical terms and the answer seems to be intuitive. On the other hand, there is no knowledge of IT terminology associated with big data.

The results showed poor understanding of problems of big data (particularly in IT) among students of economics. There is also an increase in knowledge at the later years of study, mainly in statistics and the possibility of use of big data solutions.

IV. BIG DATA AND EDUCATION

The widespread use of Internet, mobile computing and sensor networks are generating massive amounts of data available for decision making. As the Internet of Things grows the data can be collected from more sources. The increasing storage capacity and cloud storage solutions allow the data to be gathered across longer periods of time, and a

greater number of variables can be monitored, at a relatively low cost and with less effort than ever before.

This situation brings opportunities and challenges to many domains, so the educational institutions should definitely focus on this emerging field to satisfy the market demand for professionals in big data analyses, applications and technologies.

Given the broad scope of the issues covered by big data domain, higher education in economic schools should provide an overview of the possible uses of big data in

through business oriented to strictly technological ones. Five of them refer directly to the phrase 'big data' in their titles and all of them are offered by information technology or economic faculties or specializations. These technical and economical faculties also some other courses related to big data such as programming, visualization of information, social networks etc. The term 'big data' also appears in descriptions of strictly humanistic subjects, for example offered by Faculties of Polish Studies in the context of social media.

Table II. Overview of university courses that referred to big data

no.	Course name	University / faculty or institute
1	Big data	Warsaw School of Economics
2	Big data	Wrocław University of Economics / Management IT and Finance
3	Big data and business analytics	University of Zielona Góra / IT
4	Big data processing	University of Warsaw / Mathematics, Informatics and Mechanics
5	Big data processing	University of Wrocław
6	Business intelligence systems	Nicolaus Copernicus University in Toruń / Management
7	Concurrent and Distributed Programming	University of Warsaw / Mathematics, Informatics and Mechanics
8	Culture 2.0: social media	Jagiellonian University in Kraków / Polish Studies
9	Design and visualization of information	Nicolaus Copernicus University in Toruń / History
10	International Forecasting and Simulations	University of Warsaw / International Relations
11	Landscape of today's Polish Internet	Kazimierz Wielki University in Bydgoszcz / Administration and Social Science
12	Mass media	Universitas Opoliensis / Polish Studies
13	R and Big data	Warsaw University of Technology / Mathematics and Information Science
14	Social networks and multi-agent systems	University of Zielona Góra / IT
15	Social portals - the role and importance	University of Social Sciences / IT

Source: Own elaboration

business and administration institutions. The key aspect is to develop students' ability to critically evaluate the technologies, and consequences of their application. The Authors while browsing information about educational offering were attempting to identify academic courses which are related to any aspects big data.

Table II presents an overview of university courses that strictly or somehow referred to big data. The faculty name in the second column of the table provides a context for the course name. The short research on available courses was made by analyzing the online resources of the USOS systems of different universities where the detailed syllabi of the courses are available. The syllabi contain content of subjects for each classes. The criteria to choose the course as somehow related to big data were that at least one of the topics has the keyword 'big data'.

The 15 example courses have been identified among different levels of education: bachelor or master degree, post-diploma and also studies founded from Erasmus program. It is worth to notice that, the range of faculties and their specializations is quite broad from social sciences,

The presented research shows that the term 'big data' is applied in broad range of domains. The technological aspect (in the context of applied IT solutions) is mainly visible on IT oriented fields of study and in the context of analytical methods, also can be met on IT and economic studies. The business aspect of the analyzed courses is mainly considered from the social media on humanistic studies.

The social aspect of big data is also considered. It should be noted that among the identified courses commercial applications or public administration point of view are not present. We can suppose that there are some courses that deal with business data processing but this is not exposed in the course syllabi as 'big data'. However it cannot be said that the term 'big data' is used only on technical studies.

There are first attempts to create the fields of study addressed to future big data specialists. One of those efforts is the master-level study in SGH (Warsaw School of Economics). "Advanced Analytics – Big Data" has interdisciplinary character and it is placed in scientific disciplines: economics, management sciences, mathematics and informatics. Competencies of the graduates are focused

on the problems of acquiring, ordering, storing and analysis of large data volumes using advanced tools. The competencies of the graduates are planned to be the focused on business and economic phenomena and processes.

The new field of study will educate the specialists in data acquiring – as well structural as unstructured – from many resources (i.e. data bases, data warehouses, internet, text files, sensor data and geolocation data) and their analysis which encompasses searching for hidden associations, knowledge extraction, creating prognostic and simulation models and interpretation of the results.

Among the specialization courses apart from detailed economics, there are some courses on data bases and data warehouses, decision rules, optimization methods, graph and network theory, data mining, text mining, data visualization and concept of big data. Moreover the large number of computer laboratories was foreseen with SAS software. [13]

It should be noted that the field of study described above is aimed at educating relatively few analytics and big data professionals. Another issue is education of many specialists of management, finances, administration and other fields, who should have general knowledge on big data, directed on the ability to use this concept.

V. CONCLUSIONS

Above considerations show the multiaspect and relative immaturity of big data concept. The concept is constantly evolving due to the very relative term ‘big’ and can be understood in many different ways. The research of popular paper articles, survey among students and research on educational curricula of Polish universities also confirm this hypothesis on multiaspect and nonuniform understanding of the big data concept. As the research shows, in spite of high recognition of big data term, its understanding among students of economic studies is mainly intuitive and associated with everything that is related to data storage and processing of large data volumes. The specific big data features poorly recognized by students and often highlighted by big data professionals are: lack or poor structuring of data and their processing for analytical purposes in a real-time.

Apart from poor understanding of the concept, the interesting aspect is the respondents’ faith in the business utility of big data solutions. The business aspect for this group is most important. Such attitude differs from the attitude represented by most of the journalists for whom the business aspect in most cases is an occasion for discussing and describing potential problems of privacy violation and its social repercussions.

The analysis of academic curricula related to big data also indicates its broad conceptual scope. Depending on the field of study various aspects are discussed. At present the students of economic studies are supposed to be future managers. Once on the labor market they will probably expect

to have information technologies and analytical methods to bring them valuable data to help them improve business performance. This can bring positive effects in the form of increased IT investments for data analysis and big data applications. In this situation the interest in big data technologies should be used as a chance to improve the level of education in this area, so the future managers could understand real capabilities offered by massive data processing. The education in this field is necessary not only for small group of data analysis specialists/data scientists but also (to some certain degree) to all the students of economic studies. Simultaneously, apart from providing technological knowledge the care should be taken to understanding the social aspect of big data. Improper use of personal data may lead to society’s reluctance and in consequence for example to establishment of legal regulations significantly limiting the applications of new technologies of massive data processing what could be harmful for the country economy.

REFERENCES

- [1] D. Boyd, K. Crawford, “Critical questions for big data” in *Information, Communication & Society*, Volume 15, Issue 5, 2012, pp. 662-679.
- [2] M. Tabakow, J. Korczak, B. Franczyk, „Big data – definicje, wyzwania i technologie informatyczne” in *Business Informatics*, 1 (31) 2014, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, pp. 138-153.
- [3] J. Wieczorkowski, P. Polak, “Big data: Three-aspect approach” in *Online Journal of Applied Knowledge Management*, Volume 2, Issue 2, 2014, International Institute for Applied Knowledge Management, pp. 182-196.
- [4] <http://blogs.msdn.com/b/microsoftenterpriseinsight/archive/2013/04/15/the-big-bang-how-the-big-data-explosion-is-changing-the-world.aspx>, 2013.
- [5] McKinsey Global Institute, “*Big data: The next frontier for innovation, competition and productivity*”, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation, 2011.
- [6] E. Dumbil, “What is big data? An introduction to the big data landscape.”, <http://radar.oreilly.com/2012/01/what-is-big-data.html>, 2012.
- [7] M. Rouse, “Big data”, <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>, 2011.
- [8] D. Laney, “*Application delivery strategies*”, META Group, <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- [9] J. Wieczorkowski, M. Dalek, „Problem przeciążenia informacyjnego a integracja systemów informatycznych”, *Zeszyty Naukowe nr 762 Ekonomiczne Problemy Usług nr 104*, Uniwersytet Szczeciński 2013, pp. 439-448.
- [10] P. Płoszajski, „Big Data: nowe źródło przewag i wzrostu firm”, *E-mentor*, nr 3 (50) / 2013, pp. 5-10.
- [11] R. Szupiluk, „*Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli data mining*”. Warszawa: Oficyna Wydawnicza SGH, 2013.
- [12] D. Jelonek C. Stepniak, T. Turek, “*Barriers in Creating Regional Business Spatial Community*” in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, pp. 1243-1250, DOI: 10.15439/2014F264
- [13] „*Program kształcenia na studiach drugiego stopnia w Szkole Głównej Handlowej w Warszawie na kierunku Analiza danych – big data*”, Szkoła Główna Handlowa w Warszawie, 2014.