ANALYTICAL THINKING AS A KEY COMPETENCE FOR OVERCOMING THE DATA SCIENCE DIVIDE

K. Rasheva-Yordanova, E. Iliev, B. Nikolova

University of Library Studies and Information Technologies (BULGARIA)

Abstract

The analytical thinking is a critical component of mental activity that enables people to solve problems quickly and effectively. It includes a methodical grading approach that allows complex issues to be fragmented into simpler and more manageable components. In data processing, analytical thinking takes a major role in activities such as: (1) processing data (2) making forecasts, visualizing and designing data (3) extracting knowledge from accumulated data and (4) making effective and efficient solutions. In the case of big data processing, the human analytical capability is a fundamental skill in making adequate decisions. The differences in the analytical data skills of human capital put the companies and the organizations in a position of inequality in their productivity. Nowadays we observe a new form emerges of inequality defined as data science divide. This article analyzes the importance of data analytics skills and analytical thinking in the professional profile of the data scientist.

Keywords: data analytics, data science, digital divide, data science divide.

1 INTRODUCTION

In the last few years, the term Data Science has become seriously significant and the data experts have turned into the most sought specialists in almost all spheres. Mikalef, et al. [11] note that data skills are probably the most sought in companies that have big data because the skills in the profile of a scientist allow companies to ask the right questions and modify data into practical vision. They conclude that software, infrastructure and data are insufficient to provide any value if personal skills and knowledge are not available to implement them.

The framework defining the knowledge, skills and competencies of data professionals is becoming an object of constant research. Numerous authors contribute to its clear definition, distinguishing the profession of data scientist from other professions such as data analyst [13], IT specialist, Information broker [15]. Another part of the researchers concentrate their efforts on defining the scope of competencies that build the profile of the modern data specialist. For example Sicular defines Data scientist as a widely-used specialist within a variety of organizations, making it difficult to provide a complete and consistent list of required skills. The author lays down the skills required for this specialist's profile including data manipulation (warehousing), data analysis, data conversion and communication skills [17]. Ismail limits the skills of data scientist to five main categories: business, statistics, machine training, communication and analysis [9].

Costa, C.et al. confirm the general assertion that Data Scientist is a multidisciplinary profile that seeks knowledge in several areas of learning. The authors add that this specialist relies heavily on the scientific way of doing things, so research experience is extremely important in shaping his/her profile [5]. Christozov D. examines the ability to address critical information as well as verifying sources and considering constraints of applied technologies as a key factor in generating useful knowledge from acquired information.

In his study, Christozov D. Brings down the competencies in working with Big Data to three main categories: (1) the ability to extract useful data from huge and diverse repositories; (2) the possibility of verifying the received data; (3) the ability to interpret (map) received data in the context of the problem and to derive useful patterns, links, or simply increase understanding of the circumstances surrounding the problem [6]. In this respect, Manieri, A.'s opinion is important in that the data scientist is an expert who has the ability to manipulate and extract knowledge and to transform them into meaningful value [10].

Based on the existing frameworks in a previous study, we have limited the professional profile of the data specialist to combine three main categories of skills, namely: hard skills, soft skills and analytical skills [15]. In this model the skills to work with different programming languages such as Python, R, Java, Ruby, Clojure, Matlab, Pig and SQL [12] we have added to hard skills. Along with them, in the

ISBN: 978-84-09-02709-5

same group, we also added the natural language processing skills (NLP), machine learning, conceptual modelling, statistical analysis, predictive modelling and testing of hypotheses and working with databases. The category soft skills comprises a big part of the non-technical skills related to communication, organizational business strategy and system architecture understanding. [9].

Along with soft and hard skills [2], the data scientist needs to be able to use sophisticated analyzes such as projection analysis, visualization and data modelling and machine training to predict what will happen in the future and make recommendations for improving the existing business process. It can be asserted that analytical skills are based on hard skills since for decision making, strategy development and experimental research data is handled obtained from other data based on some pre-processing [15].

In the presented model, each of the categories can be considered as a mutually dependent element of the other two categories. In other words, the skills of one of the categories complement or are the basis for the realization of a task based on the skills of the other two categories. This creates the conditions for the formation of differences between data professionals on the basis of the skills they possess.

This article presents an analysis of the key competencies to overcome the data science divide, emphasizing critical thinking. The article is organized in three relatively independent sections: The first section presents the data science divide as an element of the existing term Big Data Divide. The second section presents the importance of analytical thinking in data processing. The third section describes the relationship between data analitycs and the data science divide.

2 DATA SCIENCE DIVIDE IN BIG DATA DIVIDE FRAMEWORK

In the last few years big data divide has become a term describing:

- divide between “the Big Data rich” (companies and universities that can generate or purchase and store large datasets) and the “Big Data poor” (those excluded from access to the data, expertise, and processing power) [4];
- the asymmetric relationship between those who collect, store, and mine large quantities of data, and those whom data collection targets [1];
- “bigger” data will be available only to those with the resources to support the latest technology and the largest databases [1];
- divide between those who are able to extract and use un-anticipatable and inexplicable (as described above) findings and those who find their lives affected by the resulting decisions [1].

Against the background of volume, velocity, and variety [8] of data, today we are facing a new form of inequality in society caused by the lack of not only the skills, knowledge and competence to work with large data, but also by the lack of analytical, critical and logical thought by the ‘readers’ of data.

This is a private case of Big Data Divide manifesting itself at a company/institutional level. In its essence, big data divide can be presented as a divide stemming from the shortage or lack of one or several of the necessary resources to work with big data. Defining the big data divide framework, we include 3 basic components: (1) Technical Resources, (2) Access to Big Data and (3) Human Resources (fig. 1). The availability of each of these components in the conditions created for their normal functioning ensures the competitiveness of the companies and the absence or unsatisfactory functioning of any of these resources would create conditions for the formation of inequality.

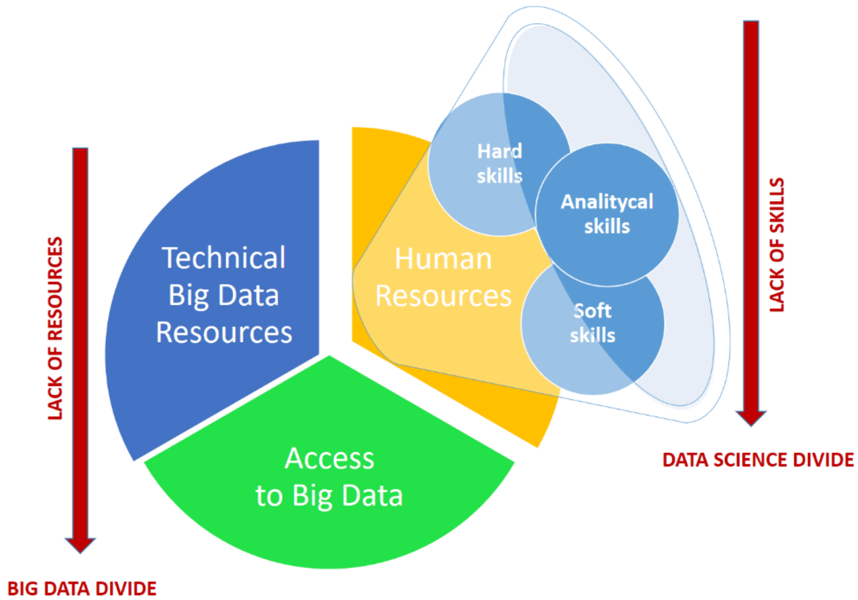

Fig. 1. Data Science Divide breakdown.

The component “Human resources“ turned out interesting to our research, the requirements of which lay the foundations for forming a subdivision of the existing big data divide. Based on the three-component model by Rasheva-Yordanova [15] it can be claimed that the human factor as a resource (Data Scientist) should possess 3 categories of skills: hard skills, soft skills and analytical skills. The lack of some of these components as well as the shortage of skills when working with big data would create conditions for the formation of a new divide we would hence call Data Science Divide.

In other words, Data Science Divide serves to identify differences in human capital and calls for a division of skill level to work with large data, to create hypotheses, to perform tests, to analyze the results, to compare with the expected values, and to make decisions based on the results obtained.

In the previous research we came across three types of data science divide [15]: (1) a division between firms with better human capital and those with no analytical skills; (2) IT specialists who can learn from big data and others who can only manage, modify and read; (3) citizens who apply analytical skills and can handle big data, drawing useful information and those who need a mediator to take advantage of the big data.

A key factor for the proper understanding, management, extraction and accumulation of knowledge in the proposed model is the analytical skills that the relevant expert possesses.

3 ANALYTICAL THINKING WHEN WORKING WITH BIG DATA

Turning data into information and knowledge can help identify critical issues, define problems and enable decision-making to be beneficial to society [16]. Analytical discoveries can lead to more efficient marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over competing organizations, and other business benefits. Generally, the main purpose of large data analyzes is to help companies make better-informed business decisions [18].

According to [19] the basic concept of data science is the extraction of useful knowledge from data to solve business problems that can be systematically treated following a process with relatively well-defined stages. The results of the scientific data require careful consideration of the context in which they will be used in the relationship between the business problem and the analysis decision. The available means of analysis can be used to find informative data elements within the big data.

The authors ... consider large data analyzes as a process of extracting meaning from big data by using specialized software systems [14]. The same authors add that analysts can use data published by data providers to analyze and generate value from them. Data scientists may have access to raw data, convert them and conduct analyses to understand the data that is useful to institutions in making decisions [14].

Data analysis requires a complex process and includes several steps such as business understanding, data collection, cleaning and pre-processing, integration, pattern recognition, analysis and interpretation of results. As with any service or artifact, price, timeliness and quality determine the success of the analysis decision [14].

It is important to note that objects similar in terms of certain features or attributes are often similar in terms of unknown attributes. Moreover, the data cannot be summed up outside the observed data to make cause-effect conclusions. It is necessary to draw attention to the available confusing factors, which often remain hidden for analysts. All of this gives us the reason to claim that when working with data, the data processing expert needs analytical skills.

In the course of working with data, however, the most important part of the analytical skills turn out to be analytical thinking skills, which allow the base material to be divided into smaller parts that make up the base material. Analytical skills in this sense allow the link between the parts and the whole material to be discovered. In the presence of critical thinking, it would be easy to identify existing problems and solve problems quickly and precisely.

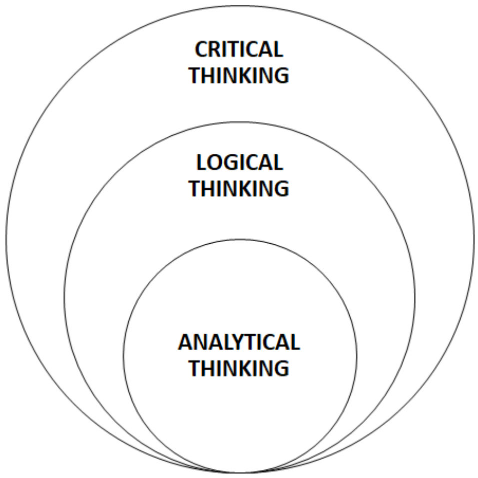

Fig. 2. Non-technical analytic skills.

All analytical skills that are technically independent and rely solely on the knowledge, experience and expertise of the Data scientist are fundamental to solving tasks to develop additional knowledge and experience. Synthesis and upgrading of skills depend primarily on analytical aspects [7].

4 LINK BETWEEN ANALYTICAL THINKING AND DATA SCIENCE DIVIDE

Part of the skills building the analytical skills in the profile of the data specialist in the proposed three category competency model “hard-soft-analytical” [15] can be:

- Comparing, classifying and sequencing abstract matter such as data and information;
- Search and detection of causal links;
- Formulation of hypotheses;
- Applying deductive and inductive thinking;
- Forecasting and planning;
- Consideration, comparison and critical assessment of alternatives;
- Generating new ideas;
- Creating original and innovative solutions;
- Modifying existing approaches;
- Recording and visualizing complex solutions.

Each of these skills is associated with a term closer to analytical thinking than to big data analysis. This gives us the reason to assert that analytical thinking is one of the skills that should be included in the data scientist’s profile. The development of knowledge and skills, respectively the expertise and experience of the data scientist, depend on the availability of analytical, critical and logical thinking.

Analytical thinking, on the other hand, can be seen as one of the key factors in overcoming the divisions that have emerged in recent years between individuals who have the ability to “read” the results of the processing of large data and the others who use the services of a mediator to benefit of the available data. In this connection, it is appropriate to make a distinction between analytical skills that are dependent on hard skills and those that are part of human thinking abilities.

In the profile of the modern data specialist, we distinguish two types of analytical skills: (1) technical analytical skills and (2) non-technical analytical skills. The first group includes all analytical skills dependent on hard skills. The process of non-technical analytical thinking includes (1) the ability to build logical chains; (2) the ability to separate the primary from the medium into a large flow of information; (3) a clear statement of thoughts and a sequence of conclusions. Therefore, the group of non-technical analytical skills consists of critical and logical thinking. The following relations are at hand:

- Analytical thinking can be supplemented with critical remarks.
- Critical thinking helps analysts to objectively consider ideas, decisions, see weaknesses and check assumptions and facts.
- Analytical thinking is closely intertwined with logical thinking and relies on this in building logical chains and relationships.
- Analytical thinking is equal to the concept of abstract-logical thinking.

Every thought-related operation is a delicate and complex process involving both internal mechanisms and external factors. Analytical thinking in combination with logical thinking helps a person to establish regularities, to forecast (calculate) the evolution of events, processes, build up possible links between objects and subjects without having to study them simultaneously, theoretically justifying the conclusions of the research with the help of the written or spoken word.

All this so far gives us the reason to assert that the application of analytical and critical thinking in performing tasks related to data processing and in particular big data will have a positive impact on the quality of the results obtained and the decisions taken. This, in turn, is one of the factors for overcoming the existing inequality in the data science divide.

5 CONCLUSIONS

Over the past few years there has been an unprecedented explosion in the organizations interest of big data, and Data Science has become a field facing a number of challenges. At a time when big data is available in all areas, the divide between specialists who can read and learn from big data and those who needs a mediator to take advantage of them becomes more and more tangible. The specificities challenges of the big data in the terms of digital divide give us reason to present the Big Data Divide as an divide existing at the resources level (technical resources, access and human resources). The digital divide in data scientist skills is presented as a special case of Big Data Divide. This form of divide we called conditionally Data Science Divide. The reasons for its generation we defined as the lack or deficiency of one of the 3 skills that builds the professional profile of the data specialist: hard skills, soft skills and analytical skills.

In this article, we consider the analytical thinking as one of the leading competencies that the data scientist should have. We believe that the knowledge and skills of the data scientiest can be adequately applied in the presence of analytical, critical and logical thinking. In turn, analytical thinking is the basis for the formation of new knowledge and skills. This gives us reason to assert that analytical thinking is a skill that should be involved in the professional profile of the data specialist, and the application of non-technical analytical skills in the learning of data would reduce the existing Data Science Divide.

ACKNOWLEDGEMENTS

This work has been supported by National Science Fund at the Ministry of Education and Science, Republic of Bulgaria, within the Project DM 12/4 - 20/12/2017.

REFERENCES

[1] Andrejevic, M. (2014, January) The Big Data Divide. International Journal of Communication 8(1):17, 1932–8036/20140005, Retrieved from http://ijoc.org/index.php/ijoc/article/view/2161/1163 on 10.05.2018.

[2] Ayankoya, K., Box, P. O., Calitz, A., Box, P. O., Greyling, J., & Box, P. O. (2014). Intrinsic Relations between Data Science, Big Data , Business Analytics and Datafication, 192–198. doi:10.1145/2664591.2664619.

[3] Begum, S., A, George. (2017). A Survey on Data Analytics Framework. International Journal of Engineering and Technology. ISSN (Print): 2319-8613. International Journal of Engineering and Technology (IJET), pp. 1650-1656.

[4] Boyd, D., Crawford, K. (2011, September). Six provocations for big data. Presentation at A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, Oxford Internet Institute, Oxford University, Oxford, UK. Available at SSRN http://ssrn.com/abstract=1926431 or http://dx.doi.org/10.2139/ssrn.1926431

[5] Costa, C., M. Y. Santos, "The data scientist profile and its representativeness in the European eCompetence framework and the skills framework for the information age," International Journal of Information Management, vol. 37, no. 6, pp. 726-734, 2017.

[6] Christozov, D., Rasheva-Yordanova K… Data Literacy: Developing Skills on Exploring Big Data Applications. International Journal of Digital Literacy and Digital Competence. Volume 8 • Issue 2 • April-June 2017. DOI: 10.4018/ IJDLDC.2017040102

[7] Ennis, Robert H, “Critical Thinking Assessment: Theory into Practice”, Vol. 32, Number 3, Summer 1993, Copyright 1993 College of Education, The OHIO State University., 2001.

[8] IBM. (2014). Bringing big data & analytics to the enterprise. Retrieved from http://www-935.ibm.com/services/multimedia/Bringing_Big_data_and_Analytics_to_the_enterprise_Workshop_for_CSIs_2014Nov13_14.pdf on 10.05.2018.

[9] Ismail, N. W. Abidin. Data Scientist Skills. IOSR Journal of Mobile Computing & Application (IOSR -JMCA) e -ISSN: 2394 - 0050, P-ISSN: 2394-0042.Volume 3, Issue 4 (Jul. -Aug. 2016), PP 52-61, DOI: 10.9790/0050-03045261

[10] Manieri, A., Demchenko, Y., Brewer, S., Hemmje, M., Riestra, R., & Frey, J. (2015). Data Science Professional uncovered How the EDISON Project will contribute to a widely accepted profile for Data Scientists. In 2015 IEEE 7th International Conference on Cloud Computing Technology and Science Data. doi:10.1109/CloudCom.2015.57

[11] Mikalef, P., I. O. Pappas, J. Krogstie, and M. Giannakos, "Big data analytics capabilities: a systematic literature review and research agenda," Information Systems and e-Business Management, pp. 1-32, 2017.

[12] Mohanty, S., M. J. and H. S. (2013). Big Data Imperatives Enterprise Big Data Warehouse, BI Implementations and Analytics. Apress.

[13] Perumal, S. (2015). Data scientist. Retrieved from http://www.slideshare.net/SevugaPerumal1/a-free- orientation-on-statistical-dataanalysis-is-conducted-on-saturday-25072015-at-10-am-and-it- has-2-hours-duration

[14] Rabhi, F. et al. (2018) Big Data Analytics Has Little to Do with Analytics. In book: Service Research and Innovation, 5th and 6th Australasian Symposium, ASSRI 2015 and ASSRI 2017, Sydney, NSW, Australia, November 2–3, 2015, and October 19–20, 2017.

[15] Rasheva-Yordanova et al. (2018) Forming of Data Science Competence for Bridging the Digital Divide. The eight edition of International Conference “Future in education” 2018. New Perspectives in Science Education, LibreriaUniversitaria Edizioni. ISSN 2384-9509 (in print).

[16] Sartika, S. Teaching Models to Increase Students' Analytical Thinking Skills. Advances in Social Science, Education and Humanities Research (ASSEHR), volume 125. 1st International Conference on Intellectuals' Global Responsibility (ICIGR 2017), Atlantis Press.

[17] Sicular, S. (2015). Big Data Analytics Failures and How to Prevent Them, 1(August).

[18] Sharmila M. et al. (2017) A Survey on Data Analytics Framework International Journal of Engineering and Technology (IJET). ISSN (Print): 2319-8613 ISSN (Online) : 0975-4024. Retrieved from http://www.enggjournals.com/ijet/docs/IJET17-09-03-010.pdf on 10.05.2018.

[19] Shum, S. B., Hall, W., Keynes, M., Baker, R. S. J., Behrens, J. T., Hawksey, M., & Jeffery, N. (2013). Educational Data Scientists : A Scarce Breed. Retrieved from http://simon.buckinghamshum.net/wp-content/uploads/2013/03/LAK13Panel-Educ Data Scientists. pdf