Formal Education in Data Science – A Perspective of Serbia

Ivan Luković [1*]
[1] University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
[*] ivan@uns.ac.rs

Abstract: *In recent years, Data Science has become an emerging education and research discipline all over the world. Software industry shows an increasing and even quite intensive interest for academic education in this area. The similar trend has been noticed in Serbia, particularly in Belgrade and Novi Sad. In this paper, we discuss main motivation factors for creating a new study program in Data Science at Faculty of Technical Sciences of University of Novi Sad. Also, we present a short survey of software industry needs for data science related experts, and discuss how we structured the new study program and addressed the main issues that come from more than evident industry requirements. The program was accredited in year 2015, both at the level of bachelor and master level studies, and this school year is its first execution, from which we expect the new experiences.*

Keywords: *Academic Education; Computing and Informatics; Data Science; Data Analytics; Information Engineering; Big Data Management*

1. INTRODUCTION

Nowadays, modern business include acquisition and store of enormous data volumes, even larger than ever before. A volume of collected data shows an enormous, practically exponential growth all over the world [4]. Most often, collected data are used in a shorter time frame, and then they are archived and almost not used, effectively. On the other hand side, such data represent a significant value that a company can utilize so as to reach created goals and provide a sustainable development. Typically, a company management is aware about significant but not exploited values ingrained in stored high data volumes. Also, we believe that the needs for generation of company knowledge form data are clearly recognized in well-matured companies. Such knowledge is to be used to raise the effectiveness of the decision and management company processes, based predominantly on quantitative, analytical methods. [1]

Unfortunately, a daily practice in many companies still intensively points out to the problem of a serious gap between the identified needs for knowledge, on one hand side, and inability of modern software products to address such needs in an effective way, on the other hand side, despite that massive data volumes already exist, while modern Information Technologies (ITs) provide the excellent technology prerequisites for a development and industry implementation of high quality software applications. We believe that this problem is just a new, "modern" form of never-ended software crisis, present for decades in many different forms, in software industry.

Some of important causes of the aforementioned problem are in the following:

a) Unsatisfactory level of organization maturity in regard to the: capacities for information management, quality management, and business processes, in general;

b) Unsatisfactory level of accumulated knowledge in a problem domain; and

c) Unsatisfactory level of accumulated knowledge in a domain of software engineering, particularly in a domain of the development and formal specification of models for software products aimed at generation of company knowledge and decision support.

In this paper, we are going to consider mainly the cause presented in c), in regard to the aspect of formal academic education, without any intention to neglect the causes a) and b), as we are quite aware that alleviating the aforementioned problem is a strategic and long life task, only possible by simultaneous addressing all its significant causes.

Following all these motives, in well-developed countries a highly emerging interest for studying a wide range of knowledge in big data processing and generation of company knowledge is present in recent years [3, 5, 6]. Such interest leads to the discipline of Data Science [2, 4]. Nowadays, a predominant attitude is that the knowledge required for development and application of Data Science is highly interdisciplinary and multidisciplinary oriented, as there is a strong need

to apply various knowledge and disciplines in a unified way in addressing the complex business problems, as a key issue in modern business [5].

Therefore, academic education in Data Science, and particularly education of engineers of such a profile, is also a strong issue in upcoming years. We notify a more intensive interest in well-developed countries for Data Science academic education even in years from 2015.

A goal of the paper is to discuss a current state, issues and development perspectives of academic education in Data Science in Serbia. Also, we present the main motivating factors for creating a new study program in Data Science at Faculty of Technical Sciences of University of Novi Sad. Also, we present a short survey of software industry needs for Data Science related experts, and discuss how we structured the new study program and addressed the main issues that come from more than evident industry requirements. The program was accredited in year 2015, both at the levels of bachelor and master studies, and the school year 2017/18 is its first execution, from which we expect the new experiences.

2. ISSUES OF ACADEMIC EDUCATION IN DATA SCIENCE

In software industry of well-developed economies, we can notify a strong fitting between the skill and education requirements for specific job positions, and the level of education and experience of software engineers or IT experts being hired at those positions. Also, such software engineers typically show a higher level of specialization to some disciplines or problem domains, as the Human Resource (HR) market is more mature. On the contrary, in under-developed or even developing economies, HR market in software industry is not mature as in well-developed economies. Therefore, fitting between the required skills and education level for the job positions and the level of education and experience of software engineers is often not appropriate, and we can notice hiring overqualified or underqualified experts at some positions, in a wider extent. The level of specialization depth of software engineers to some disciplines or problem domains is not as strong, as in well-developed economies.

However, nevertheless if we observe a software industry in well-developed or under-developed economy, the same paradox arises. More interdisciplinary oriented (software) engineers, capable of covering a wide range of tasks, knowledge and skills, are always significantly better positioned in the software industry HR market, while academic institutions motivate, often "in silence", education of more specialized experts. If we say here "more specialized experts", it means experts that are pure software

or informatics engineers, mathematicians, business administration managers, or various domain experts. On the contrary, interdisciplinary and multidisciplinary characteristics and skills are to be nurtured from early ages of academic education.

In a current academic education, we can identify study programs of the three categories, covering in some extent disciplines of Computing, Computer Science, Informatics, or Software Engineering, i.e. CSI&SE for short, as a basis to provide Data Science education. Those are: (1) Specific study programs in CSI&SE; (2) Study programs in (Applied) Mathematics; and (3) Study programs in Economics, Business Administration and Management. All large Serbian universities provide study programs from all the three categories, for many years.

Our multiyear experiences in teaching CSI&SE courses in study programs of all the three categories lead to the identification of typical students' behavioral patterns [1]. In the subsequent paragraphs, we discuss all the three patterns.

(1) Students from specific study programs in CSI&SE are predominantly technology oriented. Often, they express their animosity to the mathematical, and even more organizational, managerial or economics disciplines, with a belief that this knowledge is not necessary to them, and that someone else is to posses it. Study programs of this category often provide just a modest level of knowledge from mathematics and business administration. On the other hand side, such students express their strong interest for learning a typical technology knowledge in IT. By this, we name this behavioral pattern as "Let me learn one more technology environment, only".

(2) Students from study programs in Applied Mathematics, or just Mathematics are predominantly formally oriented. They believe that technology knowledge is of a lower level value, and also they are not aware of a necessity of having a knowledge from business administration, management, or economics. Development of skills aimed at practical application of adopted knowledge in various application domains is often underestimated or even neglected. Students from this category believe that complexity of things is just of a logical nature – the things are more complex, just if they are logically complex, while other forms of complexity are rather neglected. Study programs of this category often provide a modest level of CSI&SE knowledge, as well as business administration knowledge. By this, we name this behavioral pattern as "Let me prove one more theorem, only".

(3) Students from study programs in Economics, Business Administration and Management show a strong awareness about the importance of having

the CSI&SE and Mathematics knowledge in resolving the complex problems in organization systems. However, in a lack of their formal knowledge from these disciplines, they believe that someone else is to resolve such problems, while their task is just to rent high quality CSI&SE and Mathematics experts to resolve the problems. Study programs from this category motivate learning a highly formalized knowledge from CSI&SE and Mathematics rarely. By this, we name this behavioral pattern as "Let me follow the things globally and rent experts for strong and formal details".

Literally, we may say that the three identified behavioral patterns form "a universe of not joinable worlds". As such, a question arises (Q1) Whether such, traditional approaches to teaching selected Data Science topics can produce the appropriate experts capable of resolving complex engineering problems by the utilization of big data volumes and technologies, as well as formal modeling and data analytics methods? A derived question is (Q2) Who is capable of creating study programs that will successfully provide interfaces between all required disciplines in resolving complex organizational problems by a support of software systems?

A possible approach to addressing these questions is to profile specific study programs in the scope of CSI&SE that will nurture the appropriate level of interdisciplinarity and contribute to resolving the following two paradoxes:

(P1) More interdisciplinary oriented experts, capable of covering a wide range of tasks, knowledge and skills are always significantly better positioned in the software industry HR market, while academic institutions offer study programs that are rather self-closing, i.e. oriented to a narrower knowledge scope.

(P2) Students or young software engineers believe that they will be better positioned in software industry HR market just as they are good IT experts, i.e. programmers, while employers rather expect experts capable of recognizing and resolving their interdisciplinary oriented and complex requirements.

Following our previous teaching experiences, modern trends in many recognized world-wide universities, as well as the results obtained from the survey of software industry needs for Data Science experts already presented in [1], as well as in Sections 4 and 5 of this paper, we believe that the answer to the Q1 question is that the approaches to teaching selected Data Science topics through traditional CSI&SE, Mathematics, Economics, Business Administration and Management study programs are not enough to address the aforementioned paradoxes. As an answer to the Q2 question, we believe that profiling specific Data Science study programs in the scope of CSI&SE can successfully address the aforementioned paradoxes.

3. ACADEMIC EDUCATION IN DATA SCIENCE IN SERBIA

Apart from traditional study programs in CSI&SE in Serbia, the new initiatives for academic education in Data Science are implemented in recent years.

(A) Faculty of Technical Sciences from University of Novi Sad accredited the new study programs in Data Science in April 2015, and to the best of our knowledge, it was the earliest initiative for such study programs in Serbia and the neighboring countries, and even a very early initiative in a wider area. Those are: B.Sc. level 4-year study program in Information Engineering (240 ECTS), followed by the two M.Sc. level study programs: 1,5-year Information and Analytics Engineering (90 ECTS) and 1-year Information Engineering (60 ECTS). With an unexpected delay of two years, finally in the 2017/18 school year, Faculty of Technical Sciences initiated the first execution of the two study programs: Information Engineering at the B.Sc. and M.Sc. levels. In upcoming years, we expect initiation of the M.Sc. study program in Information and Analytics Engineering.

Apart from Information Engineering, Faculty of Technical Sciences performed in March 2018 minor changes in 1-year M.Sc. study program in Computing and Control, by including a new group of courses in High Performance Computing. The first execution of the courses in High Performance Computing is planned for the 2018/19 school year.

(B) Faculty of Sciences from University of Novi Sad accredited the new 2-year M.Sc. study module in Data Science (120 ECTS), in September 2016, in the scope of Study program in Applied Mathematics. Faculty of Sciences initiated its first execution in the 2016/17 school year. Today, it is the Study program in Applied Mathematics – Data Science, offering the two study modules: Data Analytics and High Performance Computing. To the best of our knowledge, Faculty of Sciences was the first institution initiated the execution of a Data Science study program in Serbia, as well as the first institution including the module in High Performance Computing in such program.

(C) Singidunum University from Belgrade recently accredited the new 1-year M.Sc. study program in Data Science (60 ECTS), and has opened a call for its execution in the 2018/19 year.

All the aforementioned initiatives deserve a careful attention, at least as they represent the first attempts, not only at the level of Serbia, to cover Data Science in a comprehensive way, with specific study programs. By this, Serbia has a

chance to keep up with current trends in well-developed countries, in which such initiatives are considered carefully in recent years [3, 6].

4. DATA ACQUISITION FOR THE SURVEY OF INDUSTRY NEEDS FOR DATA SCIENCE

So as to test the hypothesis about the importance of creating specific study programs in Data Science, in November 2016 we performed a short survey of the current state and needs of software companies in Serbia for the knowledge and experts in this discipline. We examined, in what extent it is already present or will be present in Serbia. [1]

The survey was based on a data acquisition process performed via anonymous web-based questionnaire that was filled in by managers from Serbian software companies. The form of anonymous questionnaire was selected to raise the level of data faithfulness in the acquisition process. We favored the anonymity criterion against the need of acquiring the exact data about respondents, their working and education profiles, as well as companies they belonged to. We believed that such a decision was not significantly violated validity of the acquired data. On the other hand side, we were aware that the potentials of acquired data for future analyses would be much greater otherwise, if we would acquired those exact data. Therefore, it is a matter of further research to repeat the same or similar survey in a couple of years, but this time by collecting exact data about respondents, and compare the results of different time snapshots and discover the patterns and trends in the area of Data Science industry applications and requirements.

With respect to the profile of a target group of respondents – often overwhelmed managers in software companies, we decided for a questionnaire of a short form that provides fill in the data in the time not longer than app. 10 minutes. The questionnaire contained 16 questions, from which 4 were in the form of a check list of already offered answers (a respondent may select as many answers as she or he likes), 9 in the form of a radio group of already offered answers (a respondent must select just one answer), while 3 were optional free text comments.

The questionnaire had three thematic groups of questions. The first group of 6 questions was related to the company profile of a respondent, profile of employed workers, implementation of projects in the area of Data Science and Business Intelligence (BI), employed workers educated in Data Science, and finally company current and future needs for hiring new experts in this discipline.

The second group consisted of 4 questions, related to the discipline of Data Science: if it is possible to acquire enough educated experts, which problem domains are the most attractive for Data Science applications, and what is a maturity level of this discipline in Serbia and in Serbia with respect to the regional countries.

The third group consisted of 5 questions, related to the academic education in Serbia: if there are specific study programs in the discipline of Data Science, what university or faculty in Serbia and abroad is favored by a respondent, if there are needs for education of specific experts in this discipline, and if there is a need for development and initiation of specific academic study programs.

The last, 16th question was a free text comment of a respondent, motivated by a need to report some additional message related to this research topic. In the data acquisition process, we collected 34 duly completed questionnaires.

5. SURVEY RESULTS AND DISCUSSION

Declared company profiles for a majority of respondents are related to the area of planning and development of software products and architectures (67,6%), as well as consultancy services in the area of software and system engineering (44,1%), or both. (The respondents could select here multiple answers.) Declared company needs for experts in CSI&SE are predominantly related to the general purpose programmers capable of participation in various IT projects (70,6%), programmers with specific knowledge in target application domains covered by companies (67,6%), while the needs for Data Science engineers are placed as 4th in order by the number of selected answers (32,4%). The respondents could also select here multiple answers. The normalized results shown in Fig. 1 better express relative ratios of various options. E.g., we notice that the requirements for Data Science engineers are approximately four times less than the requirements for programmers, as it is expectable in practice.
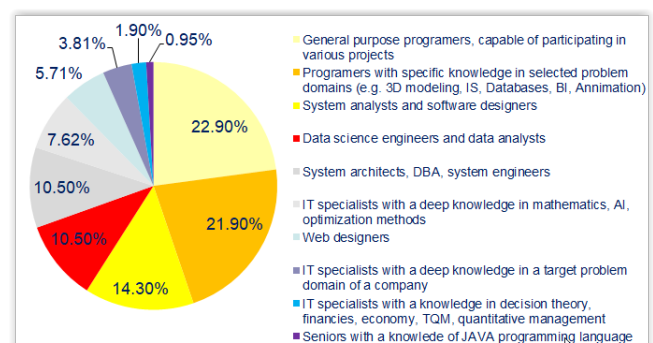

Figure 1. *Declared needs for HR in IT companies*

For two questions: Whether in your company are already implemented or will be implemented

projects in the discipline of Data Science, and then, Business Intelligence and Information Engineering, we obtained almost identical answers: about 50% of respondents declared that such projects are already implemented, or the implementation of such projects is initiated, while 29% of them declared that their companies have exact plans or clear intention to initiate such projects, while just in 24% remaining cases, such projects are not seen as close to the company goals, as shown in Fig. 2.

Figure 2. *Data Science projects in IT companies*

26.5% of respondents declared that their companies already have developed HR resources, necessary to implement the projects of such types, while 47% respondents declared clear intention or exact plans of their companies to develop such HR in the next 5 years. 26.5% of all respondents only, declared the absence of any need in their companies for HR of such type. As we asked about the existing plans in companies for employing new academic educated experts in Data Science, 67.5% of respondents declared clear needs or already existing plans that are to be implemented now or in the near future, where the number of required experts is qualified in the range from 5 to 50, as shown in Fig. 3. Only 6% respondents believe that hiring such experts is not close to the company goals, while the remaining 26.5% of them is not informed about such plans in their companies.
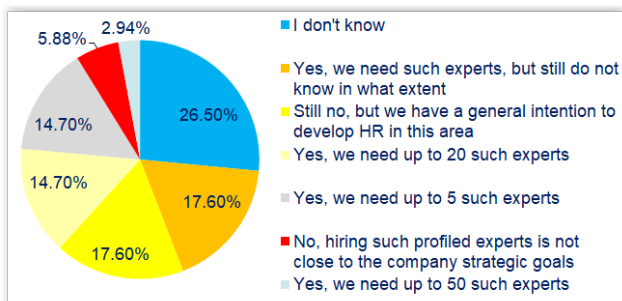

Figure 3. *Requirements for hiring Data Science experts*

44.1% respondents believe that it is not possible to acquire easily experts in Data Science, where some of them believe that such experts can be imported from abroad. However, 20.6% of them

believe that the needs for such experts significantly outperform the current HR market capacities and that their import is not a long-lasting and sustainable solution. 29.4% of respondents believe that it is possible to acquire such experts, despite that it is not easy, while there is no answer at all, committing that the number of such experts on HR market is enough to meet the company needs. See Fig. 4 for more details.

Figure 4. *Availability of Data Science experts on the HR market*

As the three most attractive application domains of Data Science, the respondents declare: a support of strategic and tactical company management in the decision process, financing, and telecommunication and Internet services. See Fig. 5 for more details.
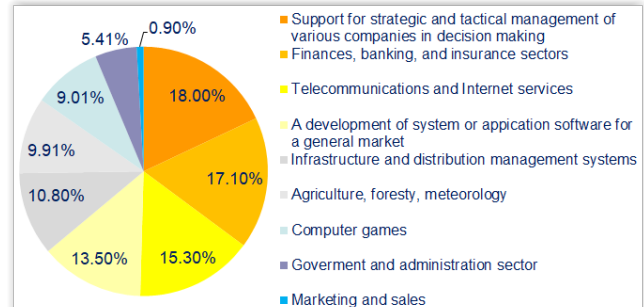

Figure 5. *The most attractive application domains of Data Science*

In regard to the maturity level of this discipline in Serbia, the most of respondents (70.6%) believe that there are many talks about Data Science, while there are still no tangible results on the market, or that the discipline is still in an early development stage, with high future expectations. The most of respondents (61.8%) are not familiar with the maturity level of this discipline in Serbia in comparison to the regional countries. The half of remaining respondents believe that Serbia is on a higher maturity level, while the other half believe that Serbia is on a lower maturity level in comparison to the regional countries.

The most of respondents believe that there are faculties in Serbia offering some of required Data Science knowledge through other study programs (40.5%), or that faculties predominantly do not

offer systematic education in this discipline, in spite of a strong need for it (35.1%), as shown in Fig. 6. A number of selections of the answer declaring that there are faculites and study programs considerably covering Data Science, or there is no need for such knowledge, and finally there are faculties offering all the required knowledge trough other, similar study programs is practically non significant.
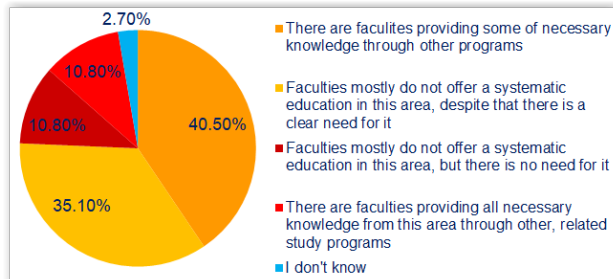
Figure 6. *Coverage of Data Science by current study programs in Serbia*

Finally, 97% respondents believe that there is a need to develop academic study programs in Data Science in Serbia.
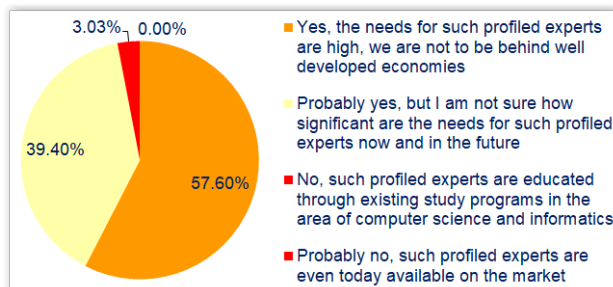
Figure 7. *Needs to develop academic study programs in Data Science in Serbia*

To summarize, we may estimate that a significant number of IT companies in Serbia, more than three quarters of all the companies, identify their strong requirements for completing Data Science projects and hiring Data Science engineers. As expected, the number of required Data Science engineers is not as huge as for general purpose software engineers. However, many companies declare their plans to hire Data Science engineers at the level of more than 5, and even up to 50 such experts. Majority of respondents notified a lack of available Data Science engineers in the HR market, and evident difficulties in acquiring the required number of such experts. They believe that current study programs in Serbia do not produce Data Science engineers having all the required knowledge, and recommend creating specific study programs in Data Science, as the needs for such expert will significantly emerge in a near future, while the discipline of Data Science is seen as one of the most promising in IT sector, in general. Since the needs for Data Science engineers are not as huge as for general purpose

software engineers, study programs in this discipline are not to be massive in terms of the maximal number of allowed students, i.e. they should be designed for a smaller number of students.

6. INFORMATION ENGINEERING PROGRAM AT FACULTY OF TECHNICAL SCIENCES

Our strategic decision in creating study programs in Data Science at Faculty of Technical Science was to cover both B.Sc. and M.Sc. academic levels. Also, we decided to design a study program for a maximum of 60 students, as the intention was not to create a massive study program. The main goals were to create a curriculum covering a body of knowledge of Data Science and Information Engineering, necessary to support information management in organization systems. The curriculum should provide a body of knowledge applicable in a wide variety of organizations of all types, or even in scientific applications. It should cover a wide range of aspects of information management, typically required by many stakeholders. It is supposed to nurture both interdisciplinary and formal approaches, where typical expected formality is at the level of mathematical rigor, whenever is possible. As Data Science is highly application oriented discipline, we decided to provide Financial Engineering, as the first application domain. Also, we left the room open for further development of various application domains.

We designed the Information Engineering study programs in a way to satisfy the following main didactic principles: a) Abstraction and Formalization, b) Quantification and Metrics, c) Specification and Implementation, and d) Communication skills. Abstraction and Formalization skills develop students' ability to understand and formalize application domain knowledge, problems and requirements, as well as to create meta-models, languages, concepts, or any kind of formalisms necessary to provide modeling of any knowledge in systems being observed. Quantification and Metrics skills develop students' ability to quantify, measure, analyze, simulate, and optimize anything that is required in any business or research, by comprehensive methods. Specification and Implementation skills develop ability to efficiently specify, develop, implement, and apply any software to address various information management requirements in business or research. Communication skills develop ability to successfully communicate and negotiate with other professionals, having different levels and range of knowledge.

Our intention is not to present here the whole study program structure, with a complete list of all designed courses. However, we just point out to the main disciplines, covered in a larger or smaller

extent by these study programs. They include the following disciplines.

Computer Science, Informatics, and Software Engineering, covering in detail all core CSI&SE disciplines according to the ACM and IEEE Computer Science curricula, and also including Formal Methods, Computational Intelligence and Machine Learning, Human-Computer Interfaces, Software Engineering, and Information Systems.

Applied Mathematics, including traditional disciplines for engineers, such as Calculus, Algebra, Numerical Calculations, and Probability Theory, as well as modern disciplines, such as Discrete Mathematics, Combinatorics, Logic, Graph Theory, Statistics, Operational Research, and Optimization Methods.

Economics, Communicology, and Management, covering the basis of Finances, Entrepreneurship in IT sector, Risks in Investment Management, Decision Theory, Business Intelligence, and Communicology, with the elements of Industrial Psychology.

General Engineering Disciplines, such us Mechanics, Time Series Processing, and Information Theory.

Apart from mandatory courses in the first two years, students can profile themselves better according to their affinities, by selecting some of elective courses in 3rd and 4th year, organized in the two main tracks: Analytical Engineering and Applied Information Engineering.

At the M.Sc. level, a pool of more than 55 elective courses is offered, covering the disciplines of Data Science, Information Engineering, High-Performance Computing, and Financial Engineering. By this, we offer many possibilities for students to profile themselves, according to their affinities, or already having jobs.

7.   CONCLUSION

Following available information on development of Data Science discipline all over the world, development of study programs at recognized universities, as well as the survey presented in this paper, we conclude that the existence of specific study programs in Data Science is a strong necessity. Faculty of Technical Sciences addressed it by the new study programs, created both at the B.Sc. and M.Sc. levels. Our first experiences collected from the program execution in its first year, show rapidly increasing interest and motivation of students for such discipline. Even the students of the first year of B.Sc. studies show a strong awareness about the importance of the Data Science discipline in their future career.

Characteristic textual comments of our respondents in a completed survey presented in this paper also clearly stress the aforementioned conclusion, and therefore we quote the three of them here. 1) "This is a unique opportunity for Serbia to become, by means of proper engineer HRs, a main exporter of such services in the region or wider, as education in the whole world is in a delay to the needs of global market, and it creates this unique opportunity for us." 2) "Those disciplines are at the early beginning in Serbia. Demands for HR are increasing, while there is a lack of high quality experts. I expect an enormous increase of market offers and demands." 3) "Analytics culture in Serbia is low and organizations do not recognize the importance of BI and how to utilize such data. It is to work on raising organization education, as today business analysis comes down just to operational reporting."

ACKNOWLEDGEMENTS

REFERENCES

[1] Luković, I., Šolaja, M. (2017) *Trends of Academic Education in Data Science – An Analysis of a Case of Serbia*. XXIII Conference Development Trends: "A Position of High Education and Science in Serbia" (TREND 2017), February 22 – 24, 2017, Zlatibor, Serbia. Proceedings, University of Novi Sad, Faculty of Technical Sciences, ISBN 978-86-7892-904-5, 162-165. (in Serbian)

[2] Smith, J. (2016) *Data Analytics: What Every Business Must Know About Big Data And Data Science*. Pinnacle Publishers, LCC, ISBN: 978-1-535-11415-8.

[3] Anderson, P., Bowring, J., McCauley, R., Pothering, G., Starr C. (2014) *An undergraduate degree in data science: curriculum and a decade of implementation experience*. Proceedings of the 45th ACM technical symposium on Computer science education (SIGCSE '14), ACM, USA, 145-150, doi: 10.1145/2538862.2538936.

[4] Provost, F., Fawcett, T. (2013) *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media, ISBN: 978-1-449-36132-7.

[5] Group of authors. (2013) *Informatics education: Europe cannot afford to miss the boat*. Report of the joint Informatics Europe & ACM Europe Working Group on Informatics Education.

[6] Golshani, F., Panchanathan, S., Friesen, O. (2000) *A Logical Foundation for an Information Engineering Curriculum*. 30th ASEE/IEEE Frontiers in Education Conference, USA, doi: 10.1109/FIE.2000.897639.