Baltic J. Modern Computing, Vol. 4 (2016), No. 3, 441-453

Data Science: Professional Requirements and Competence Evaluation

Boriss MISNEVS, Irina YATSKIV (JACKIVA)

Transport and Telecommunication Institute, Riga, Latvia

Misnevs.B@tsi.lv, Jackiva.I@tsi.lv

Abstract. The paper discusses the contemporary approach to structure of Data Science understanding in the sense of professional requirements, skills and competences for Data Science professionals. The roles of such fields as statistics, data mining, and predictive analytics are described as well as the importance of Machine Learning and Software Engineering. The list of main skills for Data Scientist is compiled. On the base of the defined skills authors suggested an approach for Data Scientist competence evaluations using Internet Portal for SE&ST Master Programs graduates competence evaluation and certification (SECEIP).

Keywords: Data Science, data analytics, requirements, Big Data, master programme, competence model.

1. Introduction

1.1. Motivation

Data Science is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured, (Dhar, 2013).

Data Science was developed to handle the flood of big data engulfing the world. A mix of statisticians, computer scientists and creative thinkers, data scientists have the following:

1. Skills to collect, to process and to extract value from giant and diverse data sets.
2. Imagination to understand, visualize and communicate their findings to non-data scientists.
3. Ability to create data-driven solutions that boost profits, reduce costs and even help save the world.

Data scientists work in every industry – from the Defence Department to Internet start-ups to financial institutions – and tackle big data projects on every level.

The McKinsey report (WEB, a) estimates that by 2018, the U.S. could face a shortage of 140,000 to 190,000 data scientists and 1.5 million managers and analysts who understand how to use big data to make decisions.

Actually the job description “Data Scientist” exists less than 10 years and the Harvard Business Review proclaimed that data scientist is the “sexiest” job of the 21st century

(WEB, b). Data scientist has become a popular job title partly because it has helped pull together a growing number of haphazardly defined and overlapping job roles.

Data science is big deal across many industries, from Retail and Finance to Travel and Transportation. In each particular field there are own major players, and different career paths.

For example, financial data specialists are capturing and analyzing new sources of data, building predictive models and running live simulations of market events. They are using technologies such as Hadoop, NoSQL and Storm to tap into non-traditional data sets (e.g., geolocation, sentiment data) and integrate them with more traditional numbers (e.g., trade data). They apply natural-language processing, text analysis and computational linguistics to source material to discover what folks really think and analyze unstructured voice recordings from call centres to recommend ways for reduction of customer churn, decreasing up-sell and cross-sell products and detect fraud.

Planes, trains and cars are equipped with a wide array of sensors that provide control centres with a continuous stream of real-time data on every aspect of the journey (e.g., driver behaviour, environment, mechanical performance, etc.). With this information, transportation data scientists are creating complex algorithms to predict and prevent the problems (Dobre and Xhafa, 2014). The Big Data comes from sensors used to gather climate, environmental information, from posts to social media sites, digital pictures and videos, purchase transaction records, or cell phone GPS signals etc. Big Data Science aims to develop new methods to store substantial amounts of data, to quickly find, analyse and validate patterns in Big Data. The number #1 in Data Science is Larry Page (CEO, Google). He has pushed the boundaries of what is possible with big data; he has accumulated the largest database on the planet.

Universities are starting to respond to the job market’s needs. And the question of the suggested program’s quality actually becomes. Universities are struggling to offer new degrees, but there’s no guarantee they’ll be worth the investment. Do you have to invest in a master’s degree? Not necessarily. Plenty of successful data scientists launched their careers with a Bachelor of Science degree and a hefty handful of skills. But this does mean that Bachelor Grade will guarantee obtaining Data Science competence.

We are strongly confirmed that only clear definition of Data Science professional requirements may be the basis for proper academic program content design and real data scientist competence evaluation. The paper is organized as follows. Next part presents some related works. Section 2 gives an analysis of Data Science requirements. In section 3 we present the competence evaluation and then we discuss the results of questionnaires, specifically designed in support of the vision of Data Sciences skills. Section 4 presents Internet Portal for SE&ST Master Programs graduates competence evaluation and certification and section 5 concludes the paper. The overall goal of this paper to review the requirements to data scientists and give the basis for Master programme outcomes definition and evaluation.

1.2. Literature review

First time the term “Data Science” was used in 1974 by Peter Nauer in (Nauer, 1974) and was defined as “the science of dealing with data, once they have been established, while the relation of the data to what they represent is delegated to other fields and sciences." Firstly the words "Data Science" were used in the title of the conference "Data science, classification, and related methods", organized in 1996 in Tokyo by the International Federation of Classification Societies. First issues of the CODATA Data Science Journal were published by the Committee on Data for Science and Technology of the International Council for Science in April 2002 (WEB, c) and Journal of Data Science Columbia University in January of 2003 (WEB, d).

The Data Science development is connected with the work of such researchers as: G.Piatetsky-Shapiro (1997), U.Fayyad (1996), D.Hand (1999, 2001), C.Aggarwal (2011), B. Mirkin (2005) etc. The fundamental concepts of data science are drawn from many fields that study data analytics. Gregory Piatetsky-Shapiro, as father of KDD and Data Mining, did a lot in popularizations of concepts, algorithms and technologies of Data Mining. 1989 G. Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop. His papers, presentations and, of course, the debates in organised seminars, active open dissemination all useful materials throw Knuggets portal develop the fundamental data science principles underlying such Data Mining (WEB, e). The paper (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) gives the following view in Knowledge Discovery in Databases (KDD) and Data Mining,: KDD refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process. Data mining is the application of specific algorithms for extracting patterns from data.

One of the famous definition suggested by Berthold and Hand (1999) "Data-mining techniques provide some useful ways to deal successfully with the sheer volume of information that constitutes one part of this problem". Data mining can be seen as the result of the natural evolution of information technologies (Han and Kamber, 2006).

In 2014 W.S.Cleveland (2014) publishes "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics" where planned "to enlarge the major areas of technical work of the field of statistics". Cleveland puts the proposed new discipline 'Data Science' in the context of computer science and the contemporary work in data mining, and suggested that the top of activities should be in multidisciplinary investigations; models and methods for data; and theory.

Data science involves principles, processes, and techniques for understanding phenomena via the (automated) analysis of data. In (Provost et.al, 2013) mentioned that the ultimate goal of data science as improving decision making and discussed the interaction between data science and business strategy, including a high-level perspective on choosing problems, to be solved with data science. The practice of data science can be described best as a combination of analytical engineering and exploration. D.Donoho in detailed review (2015) that was presented at the Turkey workshop discussed the scope of this science and how Data Science is different from Statistics.

2. Data science: professional requirements

Analysis of some core issues related to the development of common Data Science standards for higher education qualifications is needed for establishing a common way of measuring educational outcomes in the EU member countries related to actual ICT jobs requirements (Misnevs, 2015a).

Today we can find following related to Data Science career profiles: Business Analyst, Data Analyst, Data Architect, Data Engineer, Data Scientist, Marketing Analyst, Quantitative Analyst, and Statistician (WEB, f).

It is important to understand the difference in requirements for all this mentioned job positions.

For example, depending on their level of expertise, *Data Analysts* may:

1. Work with IT teams, management and/or data scientists to determine organizational goals.
2. Mine data from primary and secondary sources.
3. Clean and prune data to discard irrelevant information.
4. Analyze and interpret results using standard statistical tools and techniques.
5. Pinpoint trends, correlations and patterns in complicated data sets.
6. Identify new opportunities for process improvement.
7. Provide concise data reports and clear data visualizations for management.
8. Design, create and maintain relational databases and data systems.
9. Triage code problems and data-related issues.

Data Scientist also requires strong technical skills in:

1. Math (e.g. linear algebra, calculus and probability).
2. Statistics (e.g. hypothesis testing and summary statistics).
3. Machine learning tools and techniques (e.g. k-nearest neighbours, random forests, ensemble methods, etc.).
4. Software engineering skills (e.g. distributed computing, algorithms and data structures).
5. Data mining.
6. Data cleaning and merging.
7. Data visualization (e.g. plotting system for R ggplot2 and JavaScript library for manipulating documents based on data d3.js) and reporting techniques.
8. Unstructured data techniques.
9. R and/or SAS languages.
10. SQL databases and database querying languages.
11. Python (most common), C/C++ Java, Perl.
12. Big data platforms like Hadoop, Hive & Pig.
13. Cloud tools like Amazon S3.

Data Architect may be required to:

1. Collaborate with IT teams and management to devise a data strategy that addresses industry requirements.
2. Build an inventory of data needed to implement the architecture.
3. Research new opportunities for data acquisition.
4. Identify and evaluate current data management technologies.
5. Create a fluid, end-to-end vision for how data will flow through an organization.
6. Develop data models for database structures.
7. Design, document, construct and deploy database architectures and applications (e.g. large relational databases).
8. Integrate technical functionality (e.g. scalability, security, performance, data recovery, reliability etc.).
9. Implement measures to ensure data accuracy and accessibility.

10. Constantly monitor, refine and report on the performance of data management systems.
11. Meld new systems with existing warehouse structures.
12. Produce and enforce database development standards.
13. Maintain a corporate repository of all data architecture artefacts and procedures.

Sometimes employer is expected from one person to have all these skills. Is it possible to reach? What can we really expect from a Data Scientist? Let us provide the Definition of Data Scientist.

Data science was developed to handle the flood of big data engulfing the world. A blend of statisticians, computer scientists and creative thinkers, Data Scientists have the:

- Skills to collect process and extract value from giant and diverse data sets.
- Imagination to understand, visualize and communicate their findings to non-data scientists.
- Ability to create data-driven solutions that boost profits, reduce costs and even help save the world.

Data scientists work in every industry – from the Defence Department to Internet start-ups and financial institutions – and tackle big data projects on every level. However, as data problems become more complex, many big companies are demanding advanced degrees. Some of the industries employing Data Scientists include: Pharmaceuticals, Computer Software, Internet, Research, IT and Services, and Biotechnology.

And of course the main profile of Data Scientist must be also defined over his or her competence.

3. Data science: competence evaluation

Before discussing Data Scientists competence evaluation let us recollect definitions of Learning Outcome and Competence in general (Ibbett, 2016).

Learning Outcomes.

1. Provide a structured mechanism for describing degree programmes for the benefit of:
   (a) students
   (b) academic staff
   (c) employers
   (d) accreditation bodies
2. Are normally defined at two levels:
   (a) programme level - Programme Learning Outcomes (PLOs)
   (b) course (module) level - Intended Learning Outcomes (ILOs)
3. Should be expressed in terms of
   (a) knowledge
   (b) competences (skills),

i.e. what the student should (a) know and (b) be able to do at the end of the programme or course

Competences. Learning outcomes should be expressed using competence words. Examples are:

- state (e.g. facts);
- describe (e.g. systems);
- discuss (e.g. facts, systems, techniques);

- explain (e.g. concepts);
- compare/contrast;
- derive (e.g. equations, formulae);
- design (e.g. systems, programs);
- solve (e.g. problems, equations);
- use (e.g. tools, techniques).

By description done by Pete Skomoroch, Principal Data Scientist at LinkedIn, it is "better statisticians than most programmers & better programmers than most statisticians" (http://bit.ly/NHmRqu @peteskomoroch).

Let us compare two jobs: Data Scientist and Data Analyst (see figure 1).

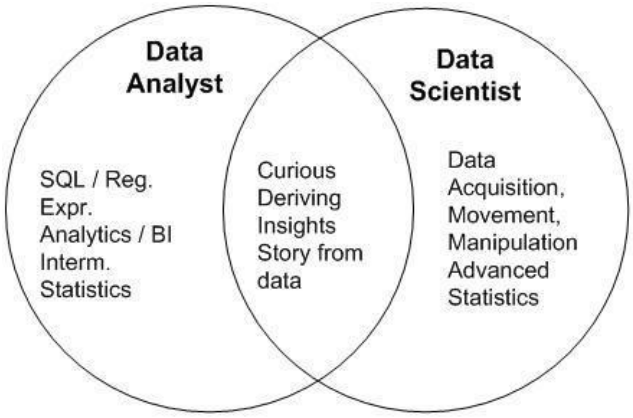

Fig. 1. Data Analyst and Data Scientist competence overlapping (WEB, g)

This Venn diagram (above) is a good first cut at describing how the two jobs overlap and how they differ. Data analysts are generally strong in SQL, they know some Regular Expressions, they can slice and dice data, they can use analytics or BI packages and they can tell a story from the data. They should also have some level of scientific curiosity.

On the other end of the spectrum, a Data Scientist will have quite a bit of machine learning and engineering or programming skills and will be able to manipulate data to his or her own will (WEB, g).

Harlan Harris et.al. (2013) suggested so called "T-Shaped Data Scientists Model". The model is presented on figure 2 and based on the following entities: Academic curiosity, Storytelling, Product sense, Statistical and Machine Learning Knowledge and Engineering experience.

A Data Scientist should have a wide breadth of abilities: academic curiosity, storytelling, product sense and engineering experience. But he or she should also have deep domain expertise in Statistical and Machine Learning Knowledge.

Academic curiosity is a desire to go beneath the surface and distil a problem into a very clear set of hypotheses that can be tested. Academic curiosity is a strong predictor of future performance. He or she will use this academic curiosity to look at the available data sets and sources to figure out an experiment or a model that solves one of the company's problems.

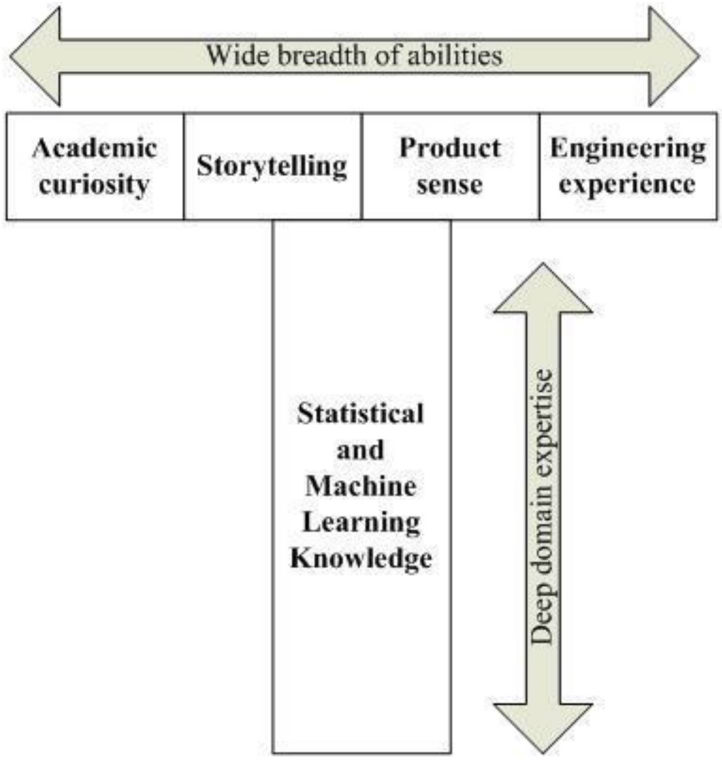

Fig. 2. T-Shaped Data Scientists Model

Storytelling is the ability to communicate your findings effectively to non technical stakeholders. This ability to distil a quantitative result from a machine learning model into something (be it words, pictures, charts, etc) that everyone can understand immediately is actually a very important skill for data scientists. As mentioned in (Cleveland, 2014) Data Scientists must be able to articulate the business value in a clear way and collaboratively work with other groups, including project sponsors and key stakeholders.

Product sense is the ability to use the story to create a new product or change an existing product in a way that improves company goals and metrics. As Data Scientist, whatever you create, in code or in algorithms, will need to translate into one of these products.

Statistical and machine learning knowledge is the domain expertise required to acquire data from different sources, create a model, optimize its accuracy, validate its purpose and confirm its significance.

The aim of supervised, machine learning is to build a model that makes predictions based on evidence in the presence of uncertainty. As adaptive algorithms identify patterns in data, a computer "learns" from the observations. When exposed to more observations, the computer improves its predictive performance.

The basic workflow (see figure 3) for obtaining a predictor model will contain the following steps (WEB, h):

4. Prepare Data
5. Choose an Algorithm
6. Fit a Model
7. Choose a Validation Method
8. Examine Fit and Update Until Satisfied
9. Use Fitted Model for Predictions

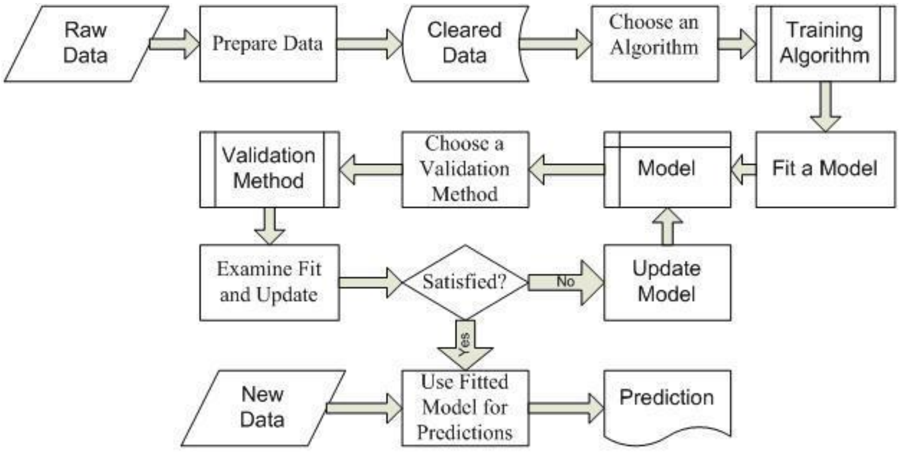

Fig. 3. Basic Machine Learning Workflow

Engineering Experience refers to the coding stuff necessary to implement and execute statistical models.

For a lot of big companies this means knowing intense amounts of Scala, Java, Closure, etc. to deploy your models into production. For start-ups this can be as simple as implementing a model in R.

Consequently, R is a great language for scaffolding models and visualization, but it's not so great for writing production ready code – it breaks whenever you throw anything more than 10 megabytes in front of it.

But, it's a great language to set up a proof of concept, and the ability to create something out of nothing and to prove that it works, is a skill that I think most data scientists ought to have.

Data Scientist must be creative to do all these things on a deadline or on constrained resources.

The difference between research scientists in academia and Data Scientists in the real world is that scientists in academia (given funding) have time to figure out problems and research a suggested solution. The general goal of academia is to move the boundary of knowledge forward at all cost. Considering themselves as academic scientists the paper authors are trying not only review the actual situation in academy but also develop an approach and tools for contemporary Data Science program development and graduates quality evaluation.

The goal of a Data Scientist in a start-up or a tech company is to move the product forward at minimal cost, yesterday. So the ability to take on deadlines, constrained resources – even your company's political climate – and push a product out in a reasonable amount of time is a really important skill.

To compare requirements for Data Scientist from academic experts and IT industry professionals we have performed an on-line survey.

Here is the list of skills we have borrowed from (Harris et.al., 2013) and asked our respondents to sort:

1. Big and Distributed Data (ex: Hadoop, Map/Reduce)
2. Business (ex: management, business development, budgeting)
3. Classical Statistics (ex: general linear model, ANOVA)
4. Front-End Programming (ex: JavaScript, HTML, CSS)
5. Graphical Models (ex: social networks, Bayes networks)
6. Machine Learning (ex: decision trees, neural nets, SVM, clustering)
7. Optimization (ex: linear, integer, convex, global)
8. Product Development (ex: design, project management)
9. Simulation (ex: discrete, agent-based, continuous)
10. Spatial Statistics (ex: geographic covariates, GIS)
11. Structured Data (ex: SQL, JSON, XML)
12. Surveys and Marketing (ex: multinomial modelling)
13. Systems Administration (ex: *nix, DBA, cloud technology)
14. Temporal Statistics (ex: forecasting, time-series analysis)
15. Unstructured Data (ex: noSQL, text mining)
16. Visualization (ex: statistical graphics, mapping, web-based datavision)

Table 1. Survey Results

Place	*Academic Experts*	*IT Professionals*	*Total Result*
1	Big and Distributed Data	Big and Distributed Data	Big and Distributed Data
2	Spatial Statistics	Temporal Statistics, Machine Learning	Temporal Statistics
3	Classical Statistics	Simulation	Machine Learning
4	Machine Learning	Structured Data, Unstructured Data	Simulation
5	Front-End Programming	Systems Administration	Classical Statistics, Optimization,
6	Graphical Models, Optimization, Simulation	Classical Statistics, Visualization	Graphical Models, Spatial Statistics

The survey results are presented in the Table 1. Areas of Competence were arranged by obtained ratings for two groups of respondents: Academic Experts and IT Professionals. Totally we have surveyed 58 recipients, somehow related to Data Science, from seven European Counties (57.8% of them defined herself as Academic Expert and other 42.2% - as IT Professional).

This survey results define requirements to Data Scientist competences in Big and Distributed Data, Temporal Statistics, Machine Learning and Simulation as the most important from the list. The first place for requirements for competences in Big and Distributed Date we may considered as statistically meaningful.

We may agree with Mike Driscoll (@dataspora) that statistics is the “grammar of data science.” It is crucial to “making data speak coherently”.

As a Good Practice of Data Scientist Master Program implementation we can mention Big Data Module at University of Liverpool (WEB, i).

It takes 8 weeks of study and provides the following Learning Outcomes:

- An in-depth and critical understanding of the concept of ‘Big Data’, the analytic techniques that can be used with respect to Big Data, and how these techniques can be used to gain competitive advantages
- An awareness of major cases of Big Data usage in science and industry and the associated Big Data challenges
- A critical understanding of the Big Data Architecture Framework, the main components and their inter-relation
- A critical understanding of Big Data infrastructures
- A critical understanding of data structures used in the context of Big Data
- An understanding of Big Data Analytics tools and platforms and methods
- An understanding of Big Data security and protection issues
- A critical awareness of the commercial relevance of Big Data.

4. Competence evaluations using Internet portal

On the base of the defined skills authors suggested an approach for Data Scientist competence evaluations using Internet Portal for SE&ST Master Programs graduates competence evaluation and certification (SECEIP) (Mishnevs, 2015a). This portal will be an outcome of the ERASMUS+ iSECRET Project implemented during 2015-2017.

The main activities the project are the followings:

- Research of European experience in SE&ST Master Programs implementation for common measurable Educational Output (competence) requirements suggestion (as a Template for Joint Master Program in SE&ST)
- Development and documenting of the Methodology for evaluation of competence in Software Engineering and Software Technologies
- Descriptions of measurable competences’ characteristics of the Master of Science for Software Engineering Program’s graduates
- Development of testing material for the Master Program Educational Outcome evaluation
- Creation of the Engineering Competence Evaluation Internet Portal (SECEIP)
- Development of on-line training course “How to use SECEIP” for academic personnel and master program graduates.

The Portal is designed as a flexible multifunctional platform for evaluation service providing. To perform the dedicated role the Portal will have the following functional parts:

1. The Master Program structure (subjects list and subjects' relationship)
2. Program's Educational Outcome
3. Subjects' Learning Outcomes
4. Information Section (guidance on the use of tests for verification purposes, actual learning outcomes achieved, guidance on the use and interpretation of results, etc.).
5. Administrative section (access and user registration, accounting results, addition and modification of materials from the site etc.)
6. Analytical section (grouping of results, preparation of summary reports, graphical representation of the results, statistical analysis etc.)
7. Community forum (to obtain feedback from students and academic staff).

One of the main problems for the research is the problem of actual professional IT knowledge representation for the program graduate’s competence testing purposes. As the main source of the professional knowledge for this research were used existing European standards (e.g. European Qualifications Framework) adapted for on-line Educational Outcome evaluation as well as the description of more then 20 best European master programs. IT industry job requirements were borrowed from European e-Competence Framework (WEB, j).

Methodology for remote evaluation of competences will recommend a set of Outcome models for different levels of Educational Outcome (knowledge, skills and competence) measurement and evaluation in IT professional area. The Methodology will define main requirements for evaluation planning, resource alignment, testing implementation, results reporting and evaluation results mapping.

As the model for the professional area description was selected so called knowledge mapping model. Knowledge maps are node-link representations in which skills are located in nodes and connected to other related skills through a series of labelled links (Misnevs, 2015a).

The main tool for competence evaluation will be on-line testing based on competence definition model and rubrics with requirements and grading criteria (Misnevs 2015b).

5. Conclusion

The scope and impact of Data Science will expand enormously in coming decades (Donoho, 2015). The contemporary approach to structure of Data Science understanding in the sense of professional requirements, skills and competences for Data Science professionals was provided. The roles of such fields as statistics, data mining, and predictive analytics are described as well as the importance of Machine Learning and Software Engineering. The compiled list of main skills for Data Scientist is evaluated by experts and professionals in Data Science in the sense of attracting university students. Analysis of some core issues related to the development of common Data Science standards for higher education qualifications, which is needed for establishing a common way of measuring educational outcomes in the EU member countries related to actual ICT jobs requirements, was performed. On the base of the defined skills authors suggested an approach for Data Scientist competence evaluations using Internet Portal for SE&ST Master Programs graduates competence evaluation and certification (SECEIP).

Acknowledgements

The research is part of the project "Implementation of Software Engineering Competence Remote Evaluation for Master Program Graduates (iSECRET)" run by TTI, contract No. 2015-1-LV01-KA203-013439, co-financed by EC ERASMUS+ program.

References

Berthold, M., Hand, D. (1999). *Intelligent data analysis.* Berlin: Springer.

Bramer, M. (2007). *Principles of Data Mining*. Springer-Verlag.

Cleveland, W. (2014). Data science: An action plan for expanding the technical areas of the field of statistics. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 7(6),* 414-417.

Donoho, D. (2015) *50 years of Data Science.* http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf

Dhar, V. (2013) Data science and prediction. *Communications of the ACM 56 (12): 64*. doi:10.1145/2500499

Dobre, C., Xhafa, F. (2014). Intelligent services for Big Data science. *Future Generation Computer Systems, 37*, pp.267-281.

Han, J, Kamber, M. (2001) *Data Mining: Concepts and Techniques.* Morgan Kaufmann Publishers.

Hand, D., Mannila, H. and Smyth, P. (2001) *Principles of Data Mining*. Cambridge: The MIT Press.

Harris H., Murphy S., Vaisman M. (2013) *Analyzing the Analyzers. An Introspective Survey of Data Scientists and Their Work,* O'Reilly Media.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39(11),* 27-34.

Ibbett R. (2016). *The role of learning outcomes in the EQANIE accreditation process. EQUANIE Presentation*. Available at: http://www.eqanie.eu/media/Ibbett_

Loukides, M. (2010). *What is data science*? https://www.oreilly.com/ideas/what-is-data-science

Mirkin, B. (2005). *Clustering for Data Mining: A Data Recovery Approach.* NewYork: Chapman & Hall.

Misnevs, B. (2015a) Software Engineering Competence Evaluation Portal, *Procedia Computer Science, March*, 11 – 17.

Misnevs, B. (2015b) European qualifications and e-competence frameworks application for evaluation of educational outcome, *Proceedings of International Conference «Research and Technology – Step into the Future», Riga, TTI. 12 December, 2015.*

Nauer, P. (1974). *Concise Survey of Computer Methods,* Studentlitteratur.

Patil, D. (2016). *Building Data Science Teams* http://www.oreilly.com/data/free/building-data-science-teams.csp

Piatetsky-Shapiro, G. (1997). Data mining and knowledge discovery. [S.l.]: [s.n.].

Provost, F., Fawcett, T. (2013). *Data Science For Business. Sebastopol*, CA: O'Reilly Media. Print
Tan, P., Steinbach, M. and Kumar, V. (2005). *Introduction to data mining.* Boston: Pearson Addison Wesley.
Steinbach, M., Tan, P., Kumar, V. (2005). *Data mining.* Harlow: Addison-Wesley.
WEB (a). *Big data: The next frontier for competition. McKinsey report*. http://www.mckinsey.com/features/big_data
WEB (b). *MIT Technology Review (2013).* https://www.technologyreview.com/magazine/2013/05/
WEB (c). *Data Science Journal.* http://www.jstage.jst.go.jp/browse/dsj/1/0/_
WEB (d). *Journal of Data Science: Volume 1, Number 1, January 2003.* http://www.jds-online.com/v1-1
WEB (e). *Analytics, Data Mining, and Data Science* http://www.kdnuggets.com/
WEB (f). *Master's in Data Science | Your Guide to Data Science Graduate Programs.* http://www.mastersindatascience.org/
WEB (g). *Data scientists vs data analysts: Why the distinction matters* https://www.import.io/post/data-scientists-vs-data-analysts-why-the-distinction-matters/
WEB (h). *Supervised Learning Workflow and Algorithms, MathWorks®,* http://www.mathworks.com/help/stats/supervised-learning-machine-learning-workflow-and-algorithms.html?s_tid=gn_loc_drop
WEB (i). *MSc in Information Systems Project Management Programme Structure | University of Liverpool Online.* https://www.university-liverpool-online.com/programmes/msc-in-information-systems-project-management/structure
WEB (j). *European eCompetence Framework.* http://profiletool.ecompetences.eu/

Received April 30, 2016, accepted May 23, 2016