



ADSEE - Applied Data Science Educational Ecosystem
2019-1-HR01-KA203-060984

Co-funded by the
Erasmus+ Programme
of the European Union



ADSEE project
Applied Data Science Educational Ecosystem

Intellectual output 1: Research Analysis Report

Document Reference

Project Acronym	ADSEE			
Project Number	2019-1-HR01-KA203-060984			
Project URL	adsee.eu			
Project Manager	Algebra University College			
	Name	Željka Mojzeš	E-mail	zeljka.mojzes@algebra.hr
Name of the activity	Intellectual output 1: Research Analysis Report			
Responsible Author(s)	Algebra University College (CRO) University of Amsterdam (NL) German National Library and Leibniz Information Centre for Science and Technology (GER) Faculty of Information Studies (SLO) Arctur ltd (SLO)			
Contractual Date of Delivery	05.05.2020.			
Status of the document	FINAL			

Table of contents

Introduction	4
1. General analysis of education in DS	5
1.1 Educational providers and HEI’s which supply data science education.....	5
1.2 Academic study programs related to Data Science at HEI’s	6
1.3 Ad hoc training courses and short educations (EQF level 5 or lower).....	6
2. Market analysis	7
2.1 The demand for general profile of “data scientist” on job market.....	7
2.2 Other types of occupations required that fit profile of Data Scientist.....	8
2.3 Market needs based on domain-specific characteristic at the European level in respect to relevant occupations	8
3. Analysis of current situation at HEI’s	9
3.1 Current education provisions for courses with DS elements in partner’s countries	9
3.2 Specific needs for training / cases in technical and non-technical domains	11
3.3 Services, portals, sites and applications already in use in existing DS learning objects	12

Introduction

In order to deliver a widely usable, high-quality training in data science it is necessary to uncover, on one hand, academic supply of the study programmes related to Data Science accessible at Higher Education Institutions and other education providers in partner countries and, on the other hand, to analyze the current market needs in terms of competences and experience required for Data Scientist and other occupations using DS. Since skills needs for data scientists in different industry sectors are rarely discussed - they are in effect hired as "generalists" with rather limited domain knowledge - the aim is to understand how to satisfy an increased demand for business focused and domain-specific data science skills (i.e. in tourism or digital marketing).

It provides a detailed report with an in-depth analysis of issues/challenges different stakeholders have when approaching data science:

- a) HEIs: identification of current education provision, methodologies and challenges facing technology improvements which constantly and profoundly affects data science sector. The result will be a detailed list of existing training courses / study programs in project countries offered by HEIs. For the sector comparative purposes, analysis will also take into considerations various ad hoc short training courses offered by other education providers (EQF level 5 or lower) since they nowadays represent a trend in offering data science training programs
- b) Business sector: analyse of market needs based on domain-specific characteristic in both partner countries and at the European level in respect to relevant occupations
- c) Identification of education/training needs to be provided implementing the project activities.



1. General analysis of education in DS

As part of IO1, the partners carried out a research of education offer, market and education needs with the scope to produce a detailed report with an in-depth analysis of issues/challenges different stakeholders have when approaching Data Science. General analysis of education in DS focuses on general analysis of education in data science on a local level in each partner's country and on a global level. The result is a detailed list of existing training courses and study programs offered in the partners' countries and a global overview.

The research includes the following elements (focus on application of DS topics in non-technical programs):

- identification of current education opportunities in data science
- analysis of market needs
- identification of education/training needs
- evaluation of gathered information
- identification of the most appropriate institution for the ADSEE project

1.1 Educational providers and HEI's which supply data science education

An analysis on educational providers and HEI's which supply data science education in each partner's country but on globally available educational providers as well. It was carried out in order to identify current education provision, methodologies and challenges facing technology improvements which constantly and profoundly affects data science sector and to identify the most appropriate institution for the ADSEE project. The data has been gathered from reliable sources and was subject to diligent handling.

University of Amsterdam was identified as the most appropriate institution in the Netherlands for the ADSEE project as it has 7 faculties that are active to various degree's with data science. The Dutch have a number of historic Universities, dating back to the late 16th Era. There are 17 Universities and many applied Science Universities, see: https://en.wikipedia.org/wiki/List_of_universities_in_the_Netherlands. Data Science is embedded in the National agenda and most Universities and applied Science Universities provide indepth curriculums. The table points to the home page of each University and a number of commercial providers that are used for support via social media (LinkedIn), MOOC's or simulations. Each commercial provider provides a digital badge which can be potentially back by

the relevant University authority. In Germany currently there are 816 accredited higher education programmes with data science components. All data science programmes in Germany are high quality and in line with the quality requirements of ADSEE both in terms of curriculum design and educational content. Situation in Croatia is much different as there are only 3 providers and HEI's which supply data science education and Algebra University College was identified as the most appropriate for the ADSEE project as it gives education in various segments of Data Science; professional and academic education which is aligned with scope and focus of ADSEE project. Slovenia also has limited options in educational providers and HEI's with only two available options but both of them are highly ranked institutions on a global scale making them a good fit for the ADSEE project.

1.2 Academic study programs related to Data Science at HEI's

Academic study programs related to Data Science at HEI's were also analyzed. In the Netherlands there are 14 available study programs available, some of them offering specific content e.g. for medical students, embedded as part of a course or programs with an emphasis on Applied statistics with SPSS etc. This is suggestive of specific OER material necessary to support Researchers. University of Amsterdam, Studies: Data Science and Leiden University Applied Statistics theme for data Science were identified as the most appropriate institution for the ADSEE project. In Germany, 5 HEI's offer study programs related to Data Science all of which are high quality and in line with the quality requirements of ADSEE both in terms of curriculum design and educational content. Croatia has only one HEI that offers academic study programs related to Data Science and it is the one on Algebra University College. This graduate education is giving students excellent foundation in DataScience. ADSEE project builds upon those skills helping graduate students to go deeper in specific market niche and learn how to practically apply what they have learned during their studies. Slovenia is similar to Croatia with two available academic study programs related to Data Science both of them on highly ranked institutions on a global scale making them a good fit for the ADSEE project.

1.3 Ad hoc training courses and short educations (EQF level 5 or lower)

Analysis also included ad hoc training courses and short educations (EQF level 5 or lower) like MOOCs offered by Coursera, Udacity and various other institutions.

DS Skills are in high demand in the job market in the Netherlands. Therefore, there is a vast range of short courses available. The majority of the courses reviewed are deployed at specific locations, available at a fee to the public. The analysis included only a selection of the available courses. Local means taught in part at a physical location, but available to also to International Students. It's important to stress out that Data Science is taught in specific contexts. Therefore, for non technical researchers some of the the ADSEE OER material may need to be contextualized. Out of 11 selected institution and courses in the Netherlands, Data Science for non-data scientists, Data Science management and Big data and Statistics were identified as the most appropriate institution for



the ADSEE project. German analysis showed that oftentimes DS is being learned in an informal learning setting. For this students dedicate more and more attention to MOOCs, they mostly take the usual, large platforms like Coursera or EdX, but other sources of information like wikipedia or youtube are also very popular. Croatia has a few ad hoc training courses and short educations with two certifications offered on Algebra University College - Microsoft Professional Program (MPP) in Data Science and Data Analyst Professional Program which were identified as the most appropriate for the ADSEE project. In Slovenia, there is only one local ad hoc training course but, just as in Germany large global platforms like are very popular and accessible.

2. Market analysis

As part of IO1, carried out a research on education offer, market and education needs on a local and global level with the scope to produce a detailed report with an in-depth analysis of issues/challenges different stakeholders have when approaching Data Science.

Market analysis focuses on the current market needs in terms of competences and experience required for Data Scientist and other occupations using DS.

The research includes the following elements (focus on application of DS topics in non-technical programs):

- identification of current education opportunities in data science
- analysis of market needs
- identification of education/training need
- evaluation of gathered information

2.1 The demand for general profile of “data scientist” on job market

An analysis based on a Google trends search was conducted and it was comprised of an online survey of data scientists at the popular R blogger website, and a sample of 1000 unique Dutch online job advertisements sampled from 268 online recruitment websites, advertised 3998 times. The advert sample was provided by Kai Lemkes from Data Science District and on average was placed on 4 recruitment sites. For the European Commissions perspective on skills changes by 2030 in Holland review the Skills panorama portal. A survey predicts an 11.1% growth in the IT sector employment by 2030 with processional services increasing by just under 20%. Therefore, data scientists find themselves in one of the fastest growing occupation areas. It is difficult to quantify the volume of data science jobs available within the Netherlands. With the recent disturbance and probably Economic dislocation due to the Corona virus any current estimate is likely to be incorrect. Although not directly associated with the Dutch Job market, a Global survey with 270 data science respondents from the well known R-bloggers website does provide insight into the Educational Background of data scientists. The majority of the data scientists have



obtained a masters with PhD and Degree's having roughly equal proportions. In otherwords, the Data Scientists role is on average highly skilled with statistics and mathematics, Physics or an IT background being specifically mentioned. The expectations of the data scientists where the majority of data scientists believe a masters is the right level of experience to enter the job market. Technology is changing at a rapid pace. Data science is deskilling the role of data scientists. Expectations from employers who that in general a degree or a masters is expected, and much less so a PhD. This slightly different from the data scientist survey where there was a high percentage of PhD's studied. This hints at a watering down effect where demand in the job market as presented on recruitment boards has lower expectations than what is optimal for the role of data scientist. Further study is needed to verify this possibility.

The demand for DS jobs are high, tens of thousands of jobs available for people with DS background. The usual required educational background is computer science, mathematics, physics and statistics. Some of the actual skills in high demand are: Machine Learning, Python language, Statistics, Data Visualization, Big Data, R language etc. When it comes to the competencies and experiences it depends on the individual company.

There is a specific situation in Croatia where the data scientist officially does not yet exist. Mostly, it is expected that person applying for this kind of job have knowledge base in statistics, mathematical methods, programing and data visualizations with no experience needed.

2.2 Other types of occupations required that fit profile of Data Scientist

For occupation groups Mathematicians, actuaries and statisticians, System Analysts, Application development programmers, Software engineers / analysts / software development analysts and Database development experts it is visible that most of advertised jobs from those occupations groups with DS elements are in groups System Analysts and Application development programmers. But we could say that the most common related job titles are: Big data engineer, AI specialist, Business manager and Statistician. But for example, in Croatia, on private portals currently there are most jobs advertised for programmers, and less for occupations in the rest of those occupation groups, because persons with those skills are employed as students, especially in banking sector and ICT companies so this is reason for small numbers in other showed groups. General competencies belong to the Croatian and European description of data scientist profile but in this case they are scattered in those occupation groups.

2.3 Market needs based on domain-specific characteristic at the European level in respect to relevant occupations

From a random sample of 1000 job advertisements, the majority are for IT related professions. A smaller percentage < 10% in total for specializations that our supported by data science other than IT. The ISCO classification of the job advertisements given as part of the sampled data defines the broad area of the expected skillset. For example, in Germany the most important sectors are:



automotive, energy, transportation, supply chain management and research. The demand is high, as Germany just started large scale of digitization of industry and services.

Besides the abovementioned skills domain expertise/experience is also required. For junior jobs the abovementioned DS related degrees are usually required.

Business domains that account for DS elements have roles such as data engineers, data architects, business and data analysts, MIS reporting executives, statisticians, machine learning engineers, and big data engineers for making new digital product. General profile requirements and demands for EU (ISTAT Appendix 3) data scientist occupation from point of view of competencies are: Analytical/ Computing Skills, Delivering Quality and Ethical Analysis, Process Management Skills & Creative problem solving, Solid knowledge of Python or R, Programming experience, PhD in Computer Science, Computational Biology or relevant field, Knowledge of databases etc.

3. Analysis of current situation at HEI's

As part of IO1, carried out a research on education offer, market and education needs on a local and global level with the scope to produce a detailed report with an in-depth analysis of issues/challenges different stakeholders have when approaching Data Science. Analysis of current situation at HEI's focuses on the Identification of current education provisions for courses with DS elements.

It includes the following tasks (focus on application of DS topics in non-technical programs):

- identification of current education opportunities in data science
- analysis of market needs
- identification of education/training need
- evaluation of gathered information

3.1 Current education provisions for courses with DS elements in partner's countries

There are many courses at the University of Amsterdam with an element of Data Science. These include a number that do not have data science in their title, but have elements of machine learning or statistics. Reviewing the course catalog for the University for the last six years, one sees that the number of courses with the word data in their title is a small percentage of the overall set of courses. 494 courses out of 36967 published courses, which is equivalent to 1.2% of



all courses. However, if you look at the how the number of courses are increasing, you can see a rapid growth over time.

Programmes with Data Science component at TIB's academic partner Leibniz University of Hannover (all in German): Geodäsie und Geoinformatik (Bachelor of Science), Geodäsie und Geoinformatik (Master of Science), Informatik (Bachelor of Science), Informatik (Master of Science), Informatik im Bachelorstudiengang Technical Education (Bachelor of Science), Informatik im Fächerübergreifenden Bachelor (Bachelor of Science), Technische Informatik (Bachelor of Science), Technische Informatik (Master of Science).

Algebra currently offers courses with DS elements such as Data preparation, Data Warehouses and Business and Intelligence, Quantitative methods of data processing, Machine Learning Methods, Security, privacy and ethics of digital data, Analytical techniques based on large data sets, Advanced Machine Learning Methods, Visualization and analytical software tools, Cloud Analysis, Structured Analytical Techniques, Programming in Python, Databases and their development, Data Structures and Algorithms, Probability and Statistics, Security of Information Systems. All of those courses are oriented toward preparing students to use DS elements in applied computations for purposes of analyzing large public or industry or business data sets such as finding relevant patterns in genomes especially related to some diseases such as cancer or some neurodegenerative diseases and in forensic sciences, biotechnology. In Slovenia, current FIŠ provisions are:

BA level: Introduction to Informatics, Introduction to Algorithms, Introduction to Programming, Statistics 1, Information Systems, Operating systems, Introduction to Cryptography and Spatial Geometry, Databases and data modelling, Web Programming 2, Knowledge Discovery from Data, Web and Mobile Application Development Life-Cycle, Research Practice, Work Practice, Database Administration, Fundamentals of Document Systems, Development of Solutions for Small Companies, Introduction to Modelling and Simulation of Discrete and Continuous Systems, Algorithms in Computer Science, Methods of Network Analysis,

MA level: Big Data Analysis, Biometrics, Computer Forensics, Decision models, Decision Support System, Data mining, Internet Research, Multimedia Systems, Advanced Statistical Methods, Agent Based Modelling, Quality of Data, Testing, Evaluation and Auditing of Information Systems, Modelling of decision making processes, Informatisation of Organisation, Management Information Systems.

PhD level: Data Mining and Modelling and Simulation of Logistics Systems



3.2 Specific needs for training / cases in technical and non-technical domains

Analysis of courses which are offering education in data science showed that in most countries they focus on specific business domains such as Law, Business, Science, Cybersecurity etc. Classes are based on developing critical thinking of applying knowledge in all fields of labor market that are supported with examples of best practices. Particular tools and infrastructure is mostly defined at the faculty level using network analysis and basic data analytics, such as clustering. For example, in UvA, the Science faculty has its own GPU clusters and storage. Many courses use Git which is freely available for anyone.

There are many skills and competencies students can learn through the courses being taught at the partner's institutions. These skills include the ability to point to a ML problem and provide a solution to it, familiarity with the basics of relational databases and SQL language, as well as their use in web applications, familiarity with data modelling methodology, Data collection for data analysis, Systematization of data, Design and implementation of data analysis algorithms etc.

Tools/programs and trainings analyzed in this report fit the need of the job market. The issue is not the relevance of the course, however, the volume of the courses as there is currently, in general, more demand in the market than what is supplied.

These courses of nontechnical domains are focused on organization of work for business, scientific or industry areas for problem solving with guidance of the mentor. From the technical point of view different kinds of platforms are being used for databases, Big Data, cloud computation and data visualizations. Overall, for nontechnical domains is important knowledge and expertise in technical domains that are used in different contexts of nontechnical domains in which we follow world trends in job market. There is a small cluster around programming and also cleaning of data and finally statistical methods which are realistic needs within the job market.

When looking at the possibility of students being a good fit for entry positions position of data scientist at job market, it is obvious that one course in any subject is usually insufficient for any entry position. However, these courses are enough to wake up the students curiosity and possibly to further develop the skills needed for an entry position of data scientist at job market. Students with obtained knowledge in DS develop critical thinking for problem solving in different sectors in job market, from science to industry and business and as some research shows, like study »The Future of Jobs«, there will be high demand for their expertise.

3.3 Services, portals, sites and applications already in use in existing DS learning objects

Looking into services, portals, sites and applications already in use in existing DS learning objects it is noticeable that there is a vast number of available options such as: Dreamspark with 16 Microsoft products, portals for datasets exploration such as Kaggle and other open datasets. For programming and teaching, we use Jupiter notebooks, Anaconda, Mysql, Oracle platforms, Tableau, Visual studio, Merlin, Infoeduka, A popular platform for teaching materials is Moodle More traditional services and applications are a good options as well, such as educational videos, presentations, research papers, seminars, workshops etc.

A survey with the teaching staff was also conducted within this analysis and it showed that the methodologies and technology currently used as support to teach students curricula which requires some elements of data science are Phyton, R, Jupiter notebook but »traditional« methodologies such as learning from slides and books, including some online materials. Teachers give the students small but illustrative problems that are simple to solve but give a »taste« of real data science problems. Some of the challenges while trying to use certain technology / platform / service related to data science that the teaching staff is facing are more student based, such as engagement, motivation, and personalization of content.

Even though teachers mostly have support for setting up modules required for teaching students, additional help is always welcomed for preparation of new lectures, dataset, examples. Teachers find that professional support in setting up modules that will support current curricula with data science elements (statistic information, analysis, predictive models, and visualizations) would also be useful especially in the sense of awareness of new technology, methodology and guiding principles of their usage like tutorials or research papers, providing case studies, and reusable data and methods etc.

Significant differences are visible in usage of datasets. While some teachers think that the pre-prepared datasets on internet are more than enough for teaching purposes and provide them to students directly to solve the problems for e.g. homework, some say that no specific dataset available and they find that making use of text data analysis would be helpful.

When asking the teaching staff which data science area they and their students would benefit from most of them answered that the greatest benefit would come from the domain of business understanding, analytical approach and requirements for data, predictive modeling and visualization as they are the most relevant for all labor market areas. Teachers think that they could benefit from a well-prepared course taught at an established university but they lack the time to fully devote to that. They find that continuous lifelong education related to the technologies already used and the use of new emerging methods and technologies are needed for future development in this field. Even though most skills needed for job market for today and future are covered in the courses, some teachers think that there is a lack focus in training for data scientists suggesting that there is still a lot of room for improvement.



4. Appendix:

1. IO 1 - General analysis of education in DS – full partner analysis
2. Market analysis – full partner analysis
3. Analysis of current situation at HEI's – full partner analysis