

Data Science among other Data Driven Technology Domains Revisited

Authors: Yuri Demchenko , Juan J. Cuadrado-Gallego

This paper will contribute to discussion about establishing Data Science as scientific and academic discipline.

Data Science is maturing as a scientific and technology domain, it creates a basis for new emerging technologies and data driven application domains. At the same time, continuous technology evolution imposes new challenges to modern data driven organisations in technology change management and in managing organisational human/capacity resources in related data driven technologies.

It is important to revisit Data Science relation to other technology domains that have been facilitated with the development of the data driven and data centric technologies. This will provide a basis for consistent vision for future research, technology and education development and planning.

There are multiple definitions of the Data Science discipline and technology, given in different contexts, that stress/put in the centre one of the four aspects of data analysis: Data Analytics, Data Science, Machine Learning/Deep Learning, and Artificial Intelligence (see the authors book [1] for discussion)

It is important to clarify the relation of the Data Science to other closely related scientific disciplines and technology domains such as Big Data, Artificial Intelligence, Machine Learning, and Statistics. Despite the fact that some authors may refer to historical facts of mentioning these terms 10s of years ago [2], we refer to the current data driven technologies development that made Data Science a central component of all other data related and data driven technologies development. We identify such technology fusion and consolidation took place in 2011-2013 with advents of Cloud Computing and Big Data what also aligned with the National Institute of Standards and Technologies, NIST definition of the Big Data definition in 2013 [3].

Big Data serves as a technology platform to allow the Data Science and Analytics solutions and applications to work with modern data, which are of the Big Data 3V scale: Volume, Velocity, and Variety. Big Data technology platform includes large scale computation, storage and network facilities, typically cloud based, such as Hadoop, Spark, NoSQL databases, data lakes, and others.

In the whole digital economy ecosystem, the Data Science integrates all multiple components from other scientific and technology domains to drive data intensive research and emerging digital technologies development. It is important to give Data Science definition as a scientific discipline to become a foundation for academic research and curricula development:

Data Science is a complex discipline that uses conceptual and mathematical abstractions and models, statistical methods, together with modern computational tools to obtain knowledge/derive insight from data to uncover correlations and causations in business data and support decision making in scientific research and business activity.

Data Scientist is defined as a professional practicing Data Science. Starting from the first years of the Data Science and Analytics technologies adoption there were many Data Scientist definitions proposed by practitioners in the new domain that reflected their personal professional development. The following competence areas and skills were included into Data Scientist

Powered by Ju[Ubiquity Conferences (version: 1.5.14) (<http://www.ubiquitypress.com/site/conferences/>)

The experience of the EDISON Data Science Framework (EDSF) [4] development and practical implementation supported by wide research and educational community discussions allowed us to propose an actionable definition of the Data Scientist Professional, which is based on the NIST definition and extended with organisational role of the Data Scientist [1]:

A Data Scientist is a practitioner who has sufficient competences and knowledge in the overlapping regimes of expertise in data analytics skills, domain knowledge, business needs, and programming and systems engineering expertise to manage the end-to-end scientific method process through each stage in the big data lifecycle, till the delivery of an expected scientific and business value to science or industry.

The paper will also bring to discussion important aspect of the Data Scientist definition via defining the corresponding Data Science competence framework as a basis for effective Data Science education and training. Data Science requires a multi-disciplinary approach and a new type of specialists that are capable to support all stages of the data lifecycle from data production and input to data processing and actionable results delivery, visualisation and reporting. Educated and/or trained Data Scientist is becoming a critical component of the whole data driven science and technology ecosystem. It is also important to revisit the Data Scientist Professional definition to propose/identify effective approaches to Data Science competences and skills assessment that would allow developing customisable education and training curricula that would support organisational capacity building (effective HR management) and individual career development.

The paper will introduce the EDISON Data Science Framework (EDSF) that has instrumental in defining Data Science and related professional profiles competences, body of knowledge and model curricula. Experience of using EDSF for Data Science competences assessment, customized curriculum design and effective education and career path building will be presented [5]. The EDSF and related methodology has been used for defining the Data Stewardship Professional Competence Framework and Body of Knowledge [6].

References

- [1] The Data Science Framework, A View from the EDISON Project, Editors Juan J. Cuadrado-Gallego, Yuri Demchenko, Springer Nature Switzerland AG 2020, ISBN 978-3-030-51022-0, ISBN 978-3-030-51023-7 (eBook)
- [2] David Donoho, 50 Years of Data Science, Journal of Computational and Graphical Statistics, Volume 26, 2017, Issue 4, pp 745-766, Published online: 19 Dec 2017 [online]
<https://doi.org/10.1080/10618600.2017.1384734>
- [3] NIST SP 1500-1 NIST Big Data interoperability Framework (NBDIF): Volume 1: Definitions, September 2015 [online] <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf>
(<http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf>)
- [4] EDISON Data Science Framework (EDSF). EDISON Community Initiative. [online]
<https://github.com/EDISONcommunity/EDSF> (<https://github.com/EDISONcommunity/EDSF>)
- [5] Yuri Demchenko, Mathijs Maijer, Luca Comminiello, Data Scientist Professional Revisited: Competences Definition and Assessment, Professional Development and Education Path Design, International Conference on Big Data and Education (ICBDE2021), February 3-5, 2021, London, United Kingdom
- [6] Yuri Demchenko, Lennart Stoy, Research Data Management and Data Stewardship Competences in University Curriculum, In Proc. Data Science Education (DSE), Special Session, EDUCON2021 – IEEE Global Engineering Education Conference, 21-23 April 2021, Vienna, Austria

Title: [Terms and Conditions \(/site/terms-and-conditions/\)](#) | [Privacy Policy \(/site/privacy-policy/\)](#)
General Poster Submissions (/IDW-2022/sessions/469/)

Room: **Date:** June 20, 2022

Time: 08:30 - 09:30

[Back to Programme \(/IDW-2022/programme/\)](#)