



Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

The data scientist profile and its representativeness in the European e-Competence framework and the skills framework for the information age

Carlos Costa*, Maribel Yasmina Santos

ALGORITMI Research Centre, Department of Information Systems, University of Minho, Portugal

ARTICLE INFO

Keywords:

Conceptual model
Data science
Data scientist
Knowledge
Skills

ABSTRACT

The activities in our current world are mainly supported by data-driven web applications, making extensive use of databases and data services. Such phenomenon led to the rise of Data Scientists as professionals of major relevance, which extract value from data and create state-of-the-art data artifacts that generate even more increased value. During the last years, the term Data Scientist attracted significant attention. Consequently, it is relevant to understand its origin, knowledge base and skills set, in order to adequately describe its profile and distinguish it from others like Business Analyst. This work proposes a conceptual model for the professional profile of a Data Scientist and evaluates the representativeness of this profile in two commonly recognized competences/skills frameworks in the field of Information and Communications Technology (ICT), namely in the European e-Competence (e-CF) framework and the Skills Framework for the Information Age (SFIA). The results indicate that a significant part of the knowledge base and skills set of Data Scientists are related with ICT competences/skills, including programming, machine learning and databases. The Data Scientist professional profile has an adequate representativeness in these two frameworks, but it is mainly seen as a multi-disciplinary profile, combining contributes from different areas, such as computer science, statistics and mathematics.

1. Introduction

Nowadays, we live in a world surrounded by data-driven web applications. The internet we know is built upon databases and data services. The role of a Data Scientist has never been so important, making possible the creation of data products, which acquire their value from the data itself and create more data as a result (Loukides, 2010), with even more increased value.

Data Science attracted a lot of attention during the last years (Fig. 1). Press (2013) argues that Data Science emerges as the coupling of Statistics and Computer Science. In fact, if we take a closer look at the most common related search in Trends (2017), we will find the term Computer Science. Besides, in 2009, Hal Varian, Chief Economist at Google, said that “the sexy job in the next ten years will be statisticians”, referring the relevance to understand data, to process it, to extract value from it, to visualize it and to communicate results. Consequently, it seems reasonable to assume that the knowledge base expected from a Data Scientist goes beyond the skills of a computer scientist (Cleveland, 2001), or a statistician (Cleveland, 2001; Warden, 2016), or even the coupling between these two.

The term Data Scientist as a professional profile brings us to 2008, when D.J. Patil and Jeff Hammerbacher met to share experiences about

the development of the data and analytics groups at Facebook and LinkedIn, namely when they discussed what to call the people on their teams, since the respective organizations rapidly grew up (Davenport & Patil, 2012; Patil, 2011; Press, 2013). The Data Scientist profile was then established to designate someone who works on data applications that immediately and massively impact organizations (Patil, 2011), understanding how to find answers to relevant business questions, and exploring a voluminous and diverse set of data through a scientific way of doing things (Davenport & Patil, 2012).

However, when D.J. Patil and Jeff Hammerbacher met to share experiences, the term Data Scientist was not an immediate choice. Patil (2011) states: “Business Analyst seemed too limiting... Data Analyst was a contender, but we felt that title might limit what people could do... Research scientist was a reasonable job title used by companies... However, we felt that most research scientists worked on projects that were futuristic and abstract...”. The Data Science Association (DSA, 2016) also distinguishes Data Science from Business Analytics by defining the latter as “the practice of iterative, methodical exploration of an organization’s data with emphasis on statistical analysis”, instead of the inquisitive nature of Data Science. In order to conclude the discussion Data Scientist vs Data Analyst, it becomes easier to distinguish these two professional profiles if we look at the Data Scientist role as an

* Corresponding author.

E-mail addresses: carlos.costa@dsi.uminho.pt (C. Costa), maribel@dsi.uminho.pt (M.Y. Santos).

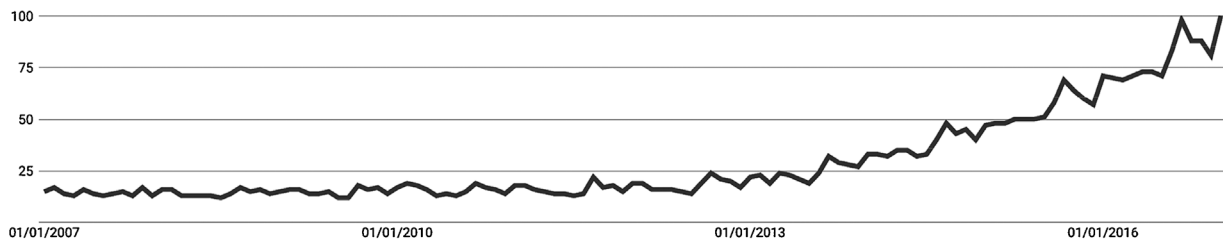


Fig. 1. Data Science's popularity in Google Trends. Retrieved from (Trends, 2017).

evolution of the Data Analyst role (IBM, 2016), regarding the ability to have a strong business acumen and communicate findings in order to surpass business challenges, namely the ones whose solution brings the most value to the organization. According to (IBM, 2016), another factor that sets these two profiles apart is the data sources, since Data Scientists tend to explore multiple disparate sources, instead of one single source. The Data Scientists have an inquisitive nature, they ask questions, in search for understanding the meaning of that data.

As previously mentioned, the knowledge base expected from a Data Scientist exceeds the one of a Computer Scientist or the one of a Statistician, it may even exceed the knowledge base that results from the combination of both profiles. Therefore, in this work, we assume that the Data Scientist profile adequately fits under the Information Systems umbrella. We do not intend to claim that the Data Scientist profile is more suitable to Information Systems than to Computer Science, Statistics, or any other field, like Experimental Physics or Systems Biology, as Davenport and Patil (2012) also identified. We simply assume that the knowledge base and skills set needed in a Data Scientist are present in the Information Systems field and we aim to understand if (and how) the Data Scientist profile is represented in the e-CF and SFIA frameworks. We made our assumption based on two relevant aspects:

1. In the framework of Bacon and Fitzgerald (2001) for the field of Information Systems, it can be observed that this field concerns automation and leveraging of “information for knowledge work, customer satisfaction and business performance” (nature of data, information and knowledge; human-computer interface; information relevance, value and cost; data quality; organizational learning...) through the use of Information and Communication Technologies (ICT), such as software, databases, data warehouses, among other concepts, which are key elements in the knowledge base and skills set of a Data Scientist, as shown in the literature and discussed later in this document;

The education map made available by the Association for Information Systems (AIS) (AIS, 2016) illustrates Data Science as an emerging main concept among 514 programs of 336 institutions. (AIS, 2016) also highlights other closely related topics, such as big data, predictive analytics, data mining, distributed systems, databases, programming and decision support systems, for example.

Regarding the ICT competences/skills frameworks analyzed in this paper, e-CF (e-CF, 2016) is described as a reference of competences required and applied at the ICT workplace, while SFIA (SFIA, 2016) is aimed towards organizational design and talent management in Information Technology (IT). Despite the disparity in terminology (ICT vs IT), these frameworks aim to standardize competences in a broad field with a relatively ambiguous use of terms. Consequently, to test the hypothesis that the Data Scientist profile is adequately represented in these frameworks, the following method was used: among the scientific community, academic formations, job opportunities, professional associations and certifications in Data Science, we gathered the knowledge base (what a Data Scientist must know about) and the skills set (what a Data Scientist is able to do). Then, a conceptual model was proposed and we look into this knowledge base and skills set in e-CF and SFIA to evaluate if the profile is adequately represented.

Summarizing, the goal of this paper is the proposal of a conceptual model for the Data Scientist profile based on the analysis of information

from the scientific community, academic formations, job opportunities, professional associations and certifications in Data Science, proposing the knowledge base (what a Data Scientist must know about) and the skills set (what a Data Scientist is able to do). This model can foster future research in this topic, and can help practitioners in recruitment campaigns for new Data Scientists or in the planning of new formations/certifications for current workers. Furthermore, this paper also aims to evaluate if e-CF and SFIA also take into consideration the Data Scientist profile as a representative profile within the ICT field of study and practice. We use the knowledge base and skills set identified in the conceptual model to evaluate if the Data Scientist profile is adequately represented in these frameworks.

This document is structured as follows: Section 2 describes related work; Section 3 summarizes the main knowledge base and skills set related to several academic programs, job opportunities, professional organizations and certifications in Data Science; Section 4 presents the proposed conceptual model; Section 5 discusses the representativeness of the Data Scientist profile in e-CF and SFIA; section 6 concludes with some remarks about the undertaken work.

2. Related work

Looking into the scientific literature, some authors already contributed to a cohesive understanding of the Data Scientist professional profile. Although this professional profile started to be emphasized somewhere in the end of the past decade (Patil, 2011), one of the first scientific publications describing what could possibly be a Data Scientist takes us as far as 2001, when Cleveland (2001) discussed the expanding of the field of statistics to embrace Data Science. Apart from that, a few other publications discuss the professional profile. Davenport and Patil (2012) present the Data Scientist as the sexiest job of the 21st century, presenting who are these people and how organizations can exploit their potential. Dhar (2013) also describes the meaning of the field and the skills present in Data Scientists. Provost and Fawcett (2013) aimed to relate the field of Data Science with other related topics, such as big data and data-driven decision making, identifying the fundamental principles present in Data Science.

The work of Loukides (2010) presents an ÓReilly Radar Report describing what is Data Science, including aspects like the data sources and its manipulation at scale, with relevant remarks concerning the description of the Data Scientist role. In Kim, Zimmermann, DeLine, & Begel, 2016, the authors discuss the emergence of Data Scientists among software engineering teams. Power (2016) looks into the core skills of Data Scientists and into the data professionals that use analytics to support the decision-making, pointing that organizations need data scientists and that academics need to prepare these professionals. The work of Kim and Lee (2016) analyzed 1240 job opportunities using a web scrapper, in order to examine how Data Scientists are seen by the industry and what is their expertise.

Although very relevant, none of the previous works proposes a conceptual model representing the profile of a Data Scientist in an organization, methodologically describing its knowledge base and its skills set based both on the academic and professional perspectives. Moreover, this work, which is an extension of Costa and Santos (2017), also evaluates the representativeness of the Data Scientist professional

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات