Investigating the Data Science Skill Gap

An Empirical Analysis

Patrick Mikalef and John Krogstie,
Department of Computer Science
Norwegian University of Science and Technology
Trondheim, Norway
e-mail: patrick.mikalef@ntnu.no

***Abstract*—With big data analytics constantly growing in importance for contemporary organizations so does the need for skilled professionals. Perhaps the most critical item noted in the age of data is the lack of people with the required skill-set to turn raw data into actionable insight. Building on this pressuring issue, the objective of this paper is to survey the status quo of technical and business-related data analytics skills in a range of different industries and identify the most important skills that will be needed in the next few years. To do so, this study builds on a sample of 202 survey responses from key executives from Norwegian firms. Our analysis reveals the level of skill-fulfilment in for technically and business-oriented employees in a number of key industries. In addition, we use survey data from an additional sample of 27 executives and interviews with 6 managers and provide a ranking of the perceived importance of data analytics-related skills according to respondents in three categories, technical skills, business and project management skills, and soft skills. Our study concludes with findings regarding the skill-gap that exists in the domain of data science as well as suggestions on how to fulfil these needs, indicating specific subject-areas that are of heightened importance.**

Keywords*—*big data analytics, data scientist, empirical, business analytics, employees

I. Introduction

In the last few years there has been an unprecedented growth in interest on use of big data analytics in the organizational context [1]. Big data analytics have been credited as enabling firms to identify previously unobtainable insight and allowing them to gain a competitive edge over their rivals by acting on this insight [2]. The value of big data analytics for the operation and survival of contemporary firms has been argued to be of increased relevance in turbulent and high-paced business environments where making informed decisions in short time-frames is critical [3]. Nevertheless, despite the large promise surrounding big data analytics, recent business reports as well as academic studies highlight that many companies and organizations are unable to leverage their investments to their full potential [4]. This has led to a stream of research that seeks to understand the key factors that underpin successful adoption and diffusion of big data analytics in the organizational context [5].

Perhaps the most widely recognized factor in enabling firms to leverage their big data analytics investments is that of human skills [6]. In fact, a prominent article by Davenport and Patil [7] makes special reference to the emerging job of the data scientist, arguing that it will become one of the most important and sought after jobs in the years to come. The article attracted much attention since it reported on findings from leading firms regarding the need for skilled people to fulfil data analytics positions and included several key implications about how educational practices need to be adapted to fulfil this growing need. Several subsequent empirical studies have confirmed the need for skilled professionals particularly in relation to technical and analytical skills [8]. Nevertheless, latest reports and studies argue that while the lack of data scientists is a well-recognized phenomenon, one of the reasons many organizations are failing to incorporate big data analytics in their operations and realize performance gains is due to the lack of skilled managers that can drive analytics projects in a top-down approach. The issue of lacking managerial skills was well described in an article by Ransbotham, et al. [9]. In their article the authors talk about the analytics gap, explaining that while companies are using increasingly more sophisticated analytical methods there is often a gap in the managers capacity to understand these new technologies and how they can be applied in business operations.

These technical and business-related data analytics skills, as well as those that are important for enabling collaborations and cooperation between experts from different domains have been argued to be critical to understand as technologies evolve and become more infused in contemporary organizations. This view is reflected by recent empirical studies such as that of Vidgen, et al. [5] and those of Gupta and George [10], Wamba, et al. [11] and Mikalef, et al. [12] where success of big data analytics initiatives is argued to be a result of multiple factors including that of technical and business-related analytics skills. Nevertheless, while we know that these types of skills are important in industry and organizations there is still limited understanding on the maturity levels throughout different industries. This is particularly crucial since domain-specific knowledge of analytics has emerged as a key consideration for many companies. Employers now are becoming increasingly more aware of the importance that domain knowledge on analytics has, seeking graduates that have these necessary skill-sets. Furthermore, while there is a growing body of research on the specific technical skills that are important for data scientists, there is still very little research on those that managers need to be proficient in, and even less on the importance of soft and interpersonal skills.

The purpose of this study is to shed some light on these issues. We start by examining the level of maturity in terms of data analytics skills for technical and business profiles. Through a survey study with 202 managers we identify the differences throughout industries identifying those that have invested more heavily in human capital and those that are lagging behind. The purpose of this analysis is to isolate those that are lacking in terms of maturity and to propose ways in which they can fulfil their skill-gaps. These findings have significant implications for study programs in the relevant fields since they can serve to provide some suggestions about modifications or additions of study modules. In sequence, we use a separate survey sample from 27 executives who report their perceptions about the importance of specific skills within the categories of technical skills, business and project management skills, and soft skills. These skills are then used to develop a ranking from the most to least important in the upcoming years. Finally, we interview 6 managers in order to identify the optimal methods for teaching such skills as well as to derive suggestions about how educational institutions should incorporate these shifts into their study programs.

II. Background

Literature and business reports have examined the range of competencies and skills that are associated with the role of the data scientist and managers of business analytics projects. For instance, Mikalef, et al. [8] explore the skill-gap that exists, and provide some details from the managers point of view regarding how companies find and recruit employees that have the necessary competencies for the jobs of the future. The authors suggest that while there is a lot of work regarding the identification of the technical skills associated with such jobs, there is little work on the managerial skills as well as on soft and collaborative skills This skill-gap is also projected to increase in the upcoming years, and particularly in relation to managerial skills on big data analytics and soft skills [13]. While there have been several studies looking into the technical skills associated with the role of the data scientist, the vast majority look into job advertisements to categorize such skill-sets [6, 14, 15]. These studies have helped to develop a better understanding of the specific skills and competences that are associated with such roles. Nevertheless, one of the limitations is that they base this classification on online job posts which may not include the full spectrum of skills. Furthermore, they look primarily at technical skills without an in-depth focus on business-related and soft skills.

The growing data economy has placed a large emphasis not only on the technical skills but also on the managerial ones. The article of Ransbotham, et al. [9] highlights the significance and importance of looking into the role of managerial skills relative to big data. The rationale is that managers will be required to be knowledgeable about increasingly more sophisticated technologies and how to apply them to meet business needs. This issue has been raised by several recent papers in which competencies such as how to define project goal in data science projects [16], being able to prioritize initiatives and link them to strategic goals [17], and having domain knowledge and foreseeing ways in which data science can help resolve business issues, are widely argued as being core aspects of contemporary managers [18]. Furthermore, there is an expanding debate around the skills that project managers should have in big data projects on coordinating and managing a big data team [18]. With big data analytics growing from a purely technical task to one that concerns the whole organization, the set of skills and competences are expanding [13]. While research has acknowledged this fact, there is still very limited knowledge about what these skills include, how they differ based on industry, and how study programs and continued education methods should be developed to fulfil them. Recent literature has highlighted the need for both technically-oriented schools as well as business schools to adapt their study programs [19], but only a very limited number of studies have actually delved into looking into the specific skills and competencies that need to be developed.

One of these studies conducted by Karbelkar and Hart [20] performed a series of semi-structured interviews with thirteen management staff of nine organizations. The authors divided noted skills into themes including good communication, prioritization of business intelligence and analytics, hunger of business intelligence, adoption for business intelligence, and other different skillsets. This study demonstrated that skills in the era of big data do not solely revolve around technical professions but also encompass managerial positions and require new forms of collaborations and communication. Looking at how these skills can be taught at higher education institutions, Rienzo and Chen [21] develop a space mapping of the processes, tools, and techniques. The authors note that what is important in the age of data is that teaching and educational curricula are dynamic and adaptable, and capable of addressing multiple levels of data analyses. In another recent study, Radovilsky, et al. [22] look at the types of skills necessary in the age of data and distinguish between competencies in technical, analytical, business and communication. The authors look at literature sources and develop a model to systemize the expected knowledge domains and skills of the fields of business data analytics and data science. They extend by looking at job-related websites and analyze the data using text mining methods.

While these studies are a starting point in recognizing, defining and delving into the whole spectrum of skills that are required in the age of data, they often rely on information posted online, therefore lacking in detail that could be provided directly by surveying those that seek after such employees. In addition, in the few studies that look into the specifics of skills, there is a lack of analysis regarding the prioritization of such skills. It is important not only to understand what new skills will emerge in the age of data, but also which ones are of the highest priority. This will enable the development of curricula to fulfill these needs and a weighted emphasis on the skills that are more critical than others. Furthermore, studies to date often overlook the differences amongst different types of industries, attributing equal importance for each [23]. It is important to identify those industries that are in greater need for analytics skills since this will also enable relevant department to adjust their study programs to accommodate such needs by developing industry-specific curricula and courses. In the following section we expand a bit further on the research questions that drive this research in relation to what has been done in past empirical research.

III. Research Questions

The breadth of knowledge required in the era of data has initiated an ongoing debate about how academic curricula should be changed to fulfil the rising needs for skilled and knowledgeable professionals [24]. Due to the increasing popularity of big data analytics in industry, several research papers argue that understanding the exact skills that are relevant in practice is of outmost importance in developing future study programs [24]. While some of the required skills are already taught in computer science, engineering, and other technological-related degrees, there are a number of competencies that are not included and highly sought after. Understanding the importance of data skills in data-oriented firms is the starting point of our study. Our research approach builds on answering the following questions:

RQ1. What is the level of data skills in different industries?

RQ2. What are the skills that are perceived as most important for the next few years?

RQ3. How can study programs and accompanying continued education be adapted to accommodate such changes and how should these skills be taught?

The first research question seeks to explore the degree to which technical and business skills related to data science are fulfilled to a dissimilar level within a number of different industries. Understanding the level of skill fulfillment in relation to industry is particularly important as it provides an indication as to which study programs need to be adjusted and include technical or business-related skills of data science. With an increasing number of industries embracing the big data paradigm there is also a growing need for professionals that have the required skill-set to leverage these novel technologies to their maximum potential. Therefore, an analysis of industries can help identify those that will have the greatest need for trained professionals in the near future. This is a particularly important question to delve into as each industry comes with a specific set of technologies and requirements around big data analytics, as well as industry-specific domain knowledge to incorporate into such curricula.

The second research question comes from a future oriented perspective and a view to develop strategies for future education. While there are a number of existing study programs that revolve around the field of big data analytics, input from industry leaders is of heightened relevance to understand what new needs will dominate industry in the next few years. High level executives have a more precise understanding of how future development will influence their industries and how these will signal new requirements in terms of education. Nevertheless, it is not only the emerging new technologies that dictate the types of skills that will be highly sought after in the near future, but also how work patterns change, and the competencies required in the evolving business landscape. For instance, while big data has resulted in a high demand in data scientists with the necessary skill-sets, it has also radically changed the ways professionals in different functional areas collaborate and work. Big data and analytics have brought together professionals from managerial, technical, and domain specific roles, resulting in a need for strong communication skills and cross-disciplinary competences. The purpose of examining the perceptions of managers for the skills that are most likely to be critical in the future is to prepare for not only the technical skill-sets that technical graduates should be taught, but also to uncover the business and interpersonal competencies that will be paramount in the unfolding new work paradigm that big data creates.

The third research question seeks to expand on the two previous ones by offering more hands-on advice about how study programs for new graduates need to be adapted, and also how continued education can be facilitated within the higher-education realm. While many organizations develop or are in the process of developing accelerated learning methods to train their existing personnel, others have limited resources to do so internally and are looking towards higher education institutions to cover these gaps. In fact, it is largely debated that higher education institutions are better equipped and more knowledgeable about best practices for teaching such skills, as well as what supporting learning technologies can be applied in each case to support efficient and effective learning outcomes. The objective of this research question is therefore to highlight the best practices that have been observed by highly experienced industry practitioners regarding optimal ways to develop study programs and methods to teach them to fulfil these new emerging needs.

IV. METHODOLOGY

A. Data Collection

To answer the research questions posited in this study, we draw on three separate samples. The first comprised of 202 industry executives that held managerial roles in IT-business operations (e.g. CIO, CTO and Digital Innovation Managers). These individuals were recruited from a contact list of Norway's 500 largest companies in terms of revenue. They were administered with a custom-built online survey and asked to evaluate the level of maturity of data analytics skills within their companies. After the initial contact, three reminder emails were sent out with a two-week internal between them. The final number of responses was 213, of which 202 were usable for further analysis. The collection of data was done over a period of approximately 4 months, between January and September 2016. Average completion time for the survey was approximately 17 minutes.

To determine the skills that would be needed in the future a group of highly experienced IT managers was contacted through the industry network of the Norwegian University of Science and technology. Following a systematic review of data-related skills, respondents were asked to indicate their perceived importance for the different skills included, and also add as many they thought were relevant and rate their importance for the upcoming years. They were handed a questionnaire where they could also add as many skills as they thought relevant and were sent 2 reminders to complete the survey following the initial contact. The final sample consisted of 27 completed responses. On average each respondent included 9 additional skills to the ones we had initially put forwarded and ranked them in importance accordingly.

Finally, a group of 6 experts was interviewed in order to determine which was the best ways to address this skill-gap and how to adapt academic curricula. The experts were interviewed using a semi-structured interview protocol which amongst other topics focused on the opinions and view of managers regarding ways in which skills ought to be taught, and the role of higher education institutions in providing such educational courses to graduates as well as to employees seeking continuing education. There was also an emphasis on the learning technologies that would be suitable for fulfilling these needs. On average, interviews lasted 68 minutes. All responses were transcribed by one of the authors and sent back to respondents to verify their accuracy. In addition, sensitive information was removed so that the interviews could not be traced back to the interviewee.

B. Sample Description

For the first quantitative sample, measures were taken to examine there was any non-response bias. We compared responding with non-responding firms in terms of industry, firm size, and revenues to make sure that no significant differences existed. We therefore performed a series of independent t-tests with the use of the software package IBM SPSS 24.0 and no significant difference was observed. In addition, we compared early responses (those received during the first 2 weeks of the study), with late responses (those received during the last two weeks of the study) in terms of performance indicators, size, and industry, and again found no statistically significant differences. These results confirmed that there were no issues of late-response bias or non-response bias either. Sample demographics of respondents of the first quantitative study are presented in Table 1 below.

Factor	Sample	Percentage
Industry		
- Bank & Financials	28	13.8%
- Consumer Goods	22	10.8%
- Oil & Gas	21	10.4%
- Industrials (Construction & Industrial goods)	19	9.4%
- ICT and Telecommunications	11	5.4%
- Technology	9	4.4%
- Media	9	4.4%
- Transport	8	3.9%
- Other (Shipping, Basic Materials, Consumer Services etc.)	75	37.1%
Size-class (in number of employees)		
- 1 – 9	1	0.5%
- 10 – 49	34	16.8%
- 50 – 249	36	17.8%
- 250+	131	64.8%
Age of Company		
- < 1 year	0	0.0%
- 1 – 4 years	5	2.4%
- 5 – 9 years	16	7.9%
- 10 – 49 years	92	45.5%
- 50+ years	89	44.0%
Job Experience (in years)		
- 0 - 10	72	35.6%
- 11 - 20	89	44.0%

- 21 - 30	36	17.8%
- 31 +	5	2.4%

Table 1 Sample demographics of first quantitative study

From the collected sample, most firms belonged to the bank and financials sector (13.8%), followed by consumer goods (10.8%), oil and gas (10.4%), and industrials (9.4%). The majority of companies, 64.8%, were large (250+ employees), 17.8% were medium-sized (50-249 employees), 16.8% small (10-49 employees), and 0.5% micro (1-9 employees). Private companies dominated the sample with 97.2%, and public-sector organizations consisting a small proportion of the sample (2.8%). The majority of respondents were highly experienced professionals, with the largest proportion of responses coming from employees with 11-20 years of experience (44.0%), and a slightly smaller percentage having between 0-10 years in the firm they operated in (35.6%). Most respondents held the position of Chief Information Officer (CIO) (36.1%), followed by head of digital strategy (20.8%), and senior vice presidents (16.3%).

With regards to the sample of firms that formed our second quantitative study, the demographics are visible in Table 2. To extract the necessary demographic information, we asked respondents to give us an answer to each respective question, and then cross-referenced their response to publicly available data. The second quantitative sample was predominantly populated by companies belonging to the technology industry (22.2%), followed by oil & gas, consumer services, consumer goods, and telecommunications, which shared an equal proportion of the sample (11.1%). Regarding size, most companies were in the large size-class (55.5%), followed by medium-sized (11.1%), small (3.7%), and micro (3.7%). The whole sample constituted of private firms, and respondents were once again highly experienced with the largest group being those that had between 21-30 years of experience (51.8%), followed by those with 0-10 years of experience (25.9%).

Factor	Sample	Percentage
Industry		
- Oil & Gas	3	11.1%
- Basic Materials	1	3.7%
- Industrials	1	3.7%
- Consumer Goods	3	11.1%
- Health Care	1	3.7%
- Consumer Services	3	11.1%
- Telecommunications	3	11.1%
- Utilities	1	3.7%
- Financials	1	3.7%
- Technology	6	22.2%
- Education	1	3.7%
- Transportation	1	3.7%
- Other	2	7.4%
Size-class (in number of employees)		
- 1 - 9	1	3.7%
- 10 - 49	1	3.7%
- 50 - 249	3	11.1%
- 250 +	22	55.5%
Sector		
- Private sector	27	100.0%
- Public sector	0	0.0%
- Non-profit organization (NPO)	0	0.0%
- Non-government organization (NGO)	0	0.0%
Job Experience (in years)		
- 0 - 10	7	25.9%
- 11 - 20	6	22.2%
- 21 - 30	14	51.8%
- 31 +	0	0.0%

Table 2 Sample demographics of second quantitative study

The final sample of interviews was performed with six key executives of private firms. The respondents held positions related to managing IT and defining the firm's digital strategy. They were highly experienced with an average of 17.3 years of prior experience in the domain and all had been working for a minimum of 6 years in the organization they were currently employed in. All companies belonged to highly competitive and dynamic industries including those of consumer services, telecommunications and transportation.

C. Measures

The survey to collect data for the first quantitative study consisted of three main parts, a) questions about the demographics of their organization, b) an assessment of the importance of data analytics in company decision making and maturity of a number of different functional areas related to big data, and c) performance measures and outcomes from big data analytics deployment. In developing the types of skills that respondents were asked to indicate their level of fulfillment within their organizations, past empirical studies were utilized as well as industry reports and commentaries on the emerging requirement for well-trained data scientists [10-12]. As a result, two main categories of skills were defined, technical and business-related data skills. Specifically, for the technical analytics skills, past empirical studies were utilized and adapted measures were created to capture the breadth of related skills [10, 11, 17]. Respondents were asked to evaluate for each of the presented skills the level to which their current personnel were proficient with them All questions were assessed in terms of a 7-point likert scale measure. Respondents were asked to evaluate how much they agreed or disagreed with a number of sentences relating to the level at which their firm has managed to secure specific skills in both technical and business-oriented competencies.

For the second quantitative study a similar approach was followed, however, it differed in the sense that a broader set of skills were evaluated as the nature of the investigation was more exploratory. To capture key emerging skills as well as ones that have been well-documented in academic literature as being important for technical, business, and interpersonal skills, a thorough analysis was conducted looking at a series of recent relevant publications [6, 8, 13-15, 20, 22, 25, 26]. Through this a list of skills was compiled falling under the three categories that were mentioned previously, technical, business, and project management. They were selected and refined based on an active discussion amongst the authors as well as through consultation with other relevant sources such as reports published by public bodies, educational institutions and popular journal publications. The skills included presented some overlap; nevertheless, the aim was to be as inclusive as possible rather than miss some important category

D. Coding

The empirical analysis was performed through an iterative process of reading, coding, and interpreting the transcribed interviews and observation notes of the 27 case studies [27]. The first of the coding process involved the identification and isolation of a large number of concepts. This was done on the basis of the thematic area which they revolved around such as educational practices, supporting technologies for learning, ongoing education methods, and methods for on-site learning. We employed an open coding scheme which allowed us to quantify the characteristics of each concept. This also enabled us to cluster primary data in a tabular structure, and through an iterative process we identified the relative concepts and notions that were related to each. The two co-authors completed the independent coding of the transcripts in accordance with defined notions of the previous sections. Each coder read the transcripts independently to identify statements that revolved around skills of big data analytics, as well as other statement of competences of human capital in general. This process was repeated until inter-rater reliability of the two coders was larger than 90 percent [28].

V. Results

A. Analysis of between-industry differences

To examine the differences between industries in terms of their skill maturity in technical and business aspects of big data analytics we employed a multi-item measurement method. Specifically, we measured the maturity of organizations in terms of their technical skills using two items. These items were based on the work of Gupta and George [10] and measured a) the level to which technical staff has the necessary skills to accomplish their job successfully, and b) the perceived level of staff training fulfilment. Similarly, for business-related skills measures were based on the work of Gupta and George [10] and asked respondents to evaluate a) the ability of analytics managers to understand the business needs of (and collaborate with) other functional managers, suppliers, and customers to determine opportunities that big data might bring to the business, b) the capacity of analytics managers to coordinate big data-related activities in ways that support other functional managers, suppliers, and customers, and c) the ability of managers to understand and evaluate the output extracted from big data. For each of the two measures, technical and business analytics skills composite measures were developed based on averages from each respondent. The results sorted by industry are depicted in Figure 1 below.

Figure 1 Skill maturity by industry and type of skill

The figure shows that there is considerable diversity concerning the level of skill fulfillment depending on industry, and also with regards to the type of skills. For instance, oils and gas, ICT and telecommunications, bank and financials, technology, and consumer goods industries appear to be at a relatively better level in both technical and managerial skills compared to other industries. On the contrary, in the education, shipping, healthcare, and offshore industries there seems to be a significant lack of both technical and managerial big data analytics skills. A striking example is the very low levels of managerial skills in the healthcare industry. This finding resonates in the industry where big data analytics is gradually becoming an increasingly important component in delivering high quality services to patients and reducing unnecessary costs. It also signals that there needs to be an initiation of efforts for developing such skills within industry and adapting curricula in higher education institutions.

B. *Skill prioritization by area*

While the first quantitative study provides an indication about the differences amongst industries and in terms of technical and managerial skills as an aggregate, it does not provide much insight concerning the specific skills that are regarded as the most important or the ones that will have the greatest demand in the next few years. To examine this issue, we conducted a ranking of the skills that according to respondents would be the most important according to their views in industry today and for the next five years. Specifically, respondents were asked to answer on a 7-point likert scale the degree to which they thought each of the presented skills was important, with 1 indicating low importance skills and 7 ones with very high importance. Table 3 presents the technical skills ranked by order of importance, from most important to least important.

Rank	Technical Skills	Value
1	Exploratory data analysis	6.29
2	Data visualization	6.22
3	Machine learning techniques	6.03
4	Data mining	5.96
5	Data modeling	5.92
6	Data cleansing/preparation	5.88
7	Research methods and empirical validation	5.88
8	Neural networks	5.63
9	Data processing	5.59
10	Statistical analysis	5.51
11	Data integration	5.48
12	Data management	5.48
13	Predictive modeling	5.21
14	Algorithms/Programming	5.14
15	Data warehousing	5.02
16	Cloud computing	4.87
17	Other artificial intelligence techniques	4.53
18	Web-based application development	4.23
19	Multiple structured programming	4.11
20	Distributed-programming and computing	3.96
21	Querying languages	3.96
22	Natural language processing	3.72

Table 3 Technical skills ordered by perceived importance

The ordering of the technical skills demonstrates that there is a strong focus on general methods of exploring and experimenting with data, visualizing outcomes, and applying different types of machine learning techniques. In addition, subjects such as data mining, modeling and cleansing rank very high in importance as they are perceived as critical components of transforming raw data into actionable insight. The results also demonstrate that there may exist multiple job roles within the technical sphere. When considering managerial skills we see a similar pattern of outcomes emerging as presented in Table 4.

Rank	Managerial Skills	Value
1	Enterprise architecture of big data	6.34
2	Big data strategy formulation	6.31
3	Understanding how to apply analytics for business problems	6.22
4	Developing big data governance practices	6.19
5	Organizational implementation plans	6.19
6	Connecting business and data science	6.11
7	Problem identification and formulation	6.03
8	Project management	6.01
9	Domain knowledge	5.96
10	Economic facts and figures	5.93
11	Digital innovation skills	5.90
12	Understanding competitive landscape and market drivers	5.78
13	Knowledge about key success factors	5.74
14	Leading big data projects	5.68
15	Orchestrating business and IT personnel	5.61
16	Developing a data-driven culture	5.43
17	Business plan development	5.31
18	Process development experience	5.25
19	Digital entrepreneurship	5.03
20	Cross-functional team management	4.97
21	Orchestrating technological and human resources	4.93
22	Marketing and sales skills	4.32

Table 4 Managerial skills ordered by perceived importance

The outcomes demonstrate that respondents see managerial skills related to big data analytics equally, and even more important than technical skills. Specifically, there seems to be a strong indication that there is a need for skills related to translating business strategy into deployment of big data analytics and managing such projects. This requirement demonstrates that there are multiple related domains of knowledge that need to be strengthened, including enterprise architecture of big data, developing and implementing big data governance practices, managing projects, and translating business problems into applied analytics. The list and ranking also underscore the importance of domain knowledge, as well as on economic facts and figures to develop robust financial measures and key performance indicators. Finally, the importance of being able to manage human capital is stressed, as being able to effectively orchestrate employees from different domains of expertise is paramount in big data analytics projects. The specific soft skills in order of importance are listed in Table 5 below.

Rank	Soft Skills	Value
1	Identifying situations requiring participative group problem solving and to utilize the proper degree and type of participation	6.12
2	Cross-disciplinary collaboration	6.08
3	Recognizing obstacles of collaborative group problem solving and implement appropriate corrective actions	6.02
4	Plan and execute work in a collective environment	5.83
5	Communication	5.81
6	Presentation skills to different disciplines	5.76
7	Work with international partners and co-workers	5.72
8	Ability to present ideas in a clear and concise manner	5.62
9	Writing clear, concise, and effective memos, reports, and documentation	5.32
10	Listen non-evaluatively and to appropriately use active listening techniques	5.28
11	Maximize consonance between verbal and non-verbal messages and to recognize and interpret non-verbal messages from others	5.12
12	Self-direction and proactiveness	5.03
13	Developing teamwork in others	4.89
14	Curiosity	4.83
15	Directing others and effectively delegating tasks	4.70
16	Coaching others and building morale	4.52
17	Teaching and activating others	4.28

Table 5 Soft skills ordered by perceived importance

The outcomes demonstrate that there are number of different soft skills that individuals working with big data should have. These include the ability to work cross-disciplinary and across geographical regions, good presentation and communication skills, as well as solid idea generation and execution competencies. The weight of importance that respondents gave to these skills indicate that they should be regarded as integral parts of educational curricula. Currently these types of skills are often overlooked, particularly in technical education programs, as they are note regarded as core competences that graduates should master. The degree of importance that respondents marked these skills is a clear sign that they need to receive more attention in formal education settings related to big data analytics.

C. *Interview outcomes*

When looking at the results from the interviews with the key executives, a number of themes emerged relevant to how such skills should be taught, the supporting technologies, methods for onsite training, and suggestions about future changes in educational curricula and continued education. Regarding the method for teaching such skills, there was agreement between all interviewees that the best way is to develop study programs that build on the areas of strategic management, decision-making, data science and information technology. The respondents argued that it is important to build on these subject areas, and not only introduce them in isolation but show theoretically and practically their relevance and applicability. There was also agreement that a mix of practical and theoretical teaching methods is the most beneficial, with an emphasis on practical examples of real-case phenomena. A proposed way of establishing this is through lectures combined with group-level projects in collaboration with industry executives.

In terms of supporting technologies, there was a suggestion by the interviewees that specific types of skills can be taught with the support of learning technologies, and particularly those of a more technical nature. Many noted that within their companies' employees were complementing their competencies though online courses provided by academic institutions (e.g. Coursera, edX etc.), video presentations (e.g. YouTube), developer websites and other online tutorials, as well as other material. One of the aspects that was noted is that these technologies have a very quick life-cycle, so most relevant information is provided online. They suggested that learning technologies that could make use of such resources, integrate them, and present them in a concise way would be more relevant than developing novel technologies from scratch.

In relation to how educational curricula need to be adjusted and how continued education could be planned for the future, respondents argued that there needs to be a stronger collaboration with industry and an ongoing interaction so that study programs are relevant. There was also mention of courses that emphasize on soft skill development, especially for employees that are already in industry. Courses that focus on collaboration, interpersonal, and presentation skills were seen as very attractive for employers, as it was noted that these are fundamental in real-business settings. The respondents also suggested that such courses be built on proven teaching approaches, and methods that make individuals more aware of their communication and collaboration patterns, as well as approaches to improve them.

VI. Discussion

The aim of this research has been to explore the changes that big data and analytics introduce in terms of working practices and the skill-sets that are necessary in this emerging context. We have attempted to answer a series of questions including, what is the difference in terms of technical and managerial skills in a number of different industries, what are the most important skill-sets that graduates should have in a number of different areas, as well as approaches and best practices in order to fulfil this skill-gap.

The research method chosen was based on a multi-method approach, including a sample of response from 202 executives in Norway, a second study of 27 responses from high level executives, and six interviewees with IT managers. The results demonstrate that there are significant differences between industries, with some struggling to find employees with the right mix of technical and business skills. These results are of increased importance for study programs in specific subject areas that are closely aligned to such industries, as well as for technical programs providing specialized degrees anchored to industries. The outcomes from our second analysis emphasized in a greater depth on the specific skills within the areas of technical, managerial, and soft skills that are emerging or of increased relevance for industries. By developing a prioritized list, we were able isolate those that should be emphasized and around which more focus should be given in higher education institutions. The results also showcase that while technical skills continue to be important in the age of big data, managerial and soft skills are of equal or greater importance [29]. With big data moving from a purely technical task to one that involved the hole organization, it is critical that there are grounded methods on linking analytics to business strategies and developing and executing such projects. Our list of skills confirms this and provide an indication of which need to be fulfilled with greater urgency. Finally, our interviews with it executives give some more information about the ways these skills should be taught, and how inter-disciplinary study programs should be developed. The role of technology-enhanced learning is linked to specific skills, as well as on the importance of new and relevant learning material.

While this study provides a more detailed look into the unfolding requirements that the big data era brings, there are several limitations that future research could seek to explore. First, there needs to be amore detailed look at the specific skills per industry. While we do provide a prioritized list of skills based on domain area, these can differ depending on the type of industry and job roles that are important in each. Second, there needs to be a further empirical assessment on the best practices of teaching each type of skill. While interviews may be a good starting point for identifying best-practices they should be assessed through research means. It is our hope that his research opens up further avenues so that the growing skill-gap in the age of big data can be minimized.

References

[1] A. McAfee, E. Brynjolfsson, and T. H. Davenport, "Big data: the management revolution," *Harvard business review,* vol. 90, no. 10, pp. 60-68, 2012.

[2] J. Manyika *et al.*, "Big data: The next frontier for innovation, competition, and productivity," 2011.

[3] T. H. Davenport, P. Barth, and R. Bean, "How big data is different," *MIT Sloan Management Review,* vol. 54, no. 1, p. 43, 2012.

[4] A. Abbasi, S. Sarker, and R. H. Chiang, "Big Data Research in Information Systems: Toward an Inclusive Research Agenda," *Journal of the Association for Information Systems,* vol. 17, no. 2, 2016.

[5] R. Vidgen, S. Shaw, and D. B. Grant, "Management challenges in creating value from business analytics," *European Journal of Operational Research,* vol. 261, no. 2, pp. 626-639, 2017.

[6] A. De Mauro, M. Greco, M. Grimaldi, and G. Nobili, "Beyond Data Scientists: a Review of Big Data Skills and Job Families," *Proceedings of IFKAD,* pp. 1844-1857, 2016.

[7] T. H. Davenport and D. Patil, "Data scientist: The Sexiest Job of the 21st Century," *Harvard business review,* vol. 90, no. 5, pp. 70-76, 2012.

[8] P. Mikalef, M. N. Giannakos, I. O. Pappas, and J. Krogstie, "The Human Side of Big Data: Understanding the skills of the data scientist in education and industry," in *Global Engineering Education Conference (EDUCON), 2018 IEEE*, Tenerife, Spain, 2018: IEEE.

[9] S. Ransbotham, D. Kiron, and P. K. Prentice, "Minding the analytics gap," *MIT Sloan Management Review,* vol. 56, no. 3, p. 63, 2015.

[10] M. Gupta and J. F. George, "Toward the development of a big data analytics capability," *Information & Management,* vol. 53, no. 8, pp. 1049-1064, 2016.

[11] S. F. Wamba, A. Gunasekaran, S. Akter, S. J.-f. Ren, R. Dubey, and S. J. Childe, "Big data analytics and firm performance: Effects of dynamic capabilities," *Journal of Business Research,* vol. 70, pp. 356-365, 2017.

[12] P. Mikalef, I. O. Pappas, J. Krogstie, and M. Giannakos, "Big data analytics capabilities: a systematic literature review and research agenda," *Information Systems and e-Business Management*, vol. 16, pp. 1-32, 2018.
[13] European Commission. (2017, 31/10/2017). *Digital Skills and Jobs Coalition*. Available: https://ec.europa.eu/digital-single-market/en/digital-skills-jobs-coalition
[14] A. De Mauro, M. Greco, M. Grimaldi, and P. Ritala, "Human resources for Big Data professions: A systematic classification of job roles and required skill sets," *Information Processing & Management*, vol. 54, no. 5, pp. 807-817, 2018.
[15] A. D. A. Stanton and W. W. Stanton, "The relationship between Big Data, data science, digital analytics and the skills and abilities needed to optimise marketing decisions," *Applied Marketing Analytics*, vol. 2, no. 3, pp. 265-279, 2016.
[16] J. Saltz, I. Shamshurin, and C. Connors, "Predicting data science sociotechnical execution challenges by categorizing data science projects," *Journal of the Association for Information Science and Technology*, vol. 68, no. 12, pp. 2720-2728, 2017.
[17] P. Mikalef, V. A. Framnes, F. Danielsen, J. Krogstie, and D. H. Olsen, "Big Data Analytics Capability: Antecedents and Business Value," in *Pacific Asia Conference on Information Systems*, Langkawi, Malaysia, 2017, 2017.
[18] J. Saltz, I. Shamshurin, and C. Connors, "A Framework for Describing Big Data Projects," in *International Conference on Business Information Systems*, 2016, pp. 183-195: Springer.
[19] N. Evangelopoulos, J. W. Clark, and S. Balkan, "Introduction: Pedagogy in Analytics and Data Science," in *Analytics and Data Science*: Springer, 2018, pp. 223-226.
[20] Y.-A. Karbelkar and M. Hart, "Skills and Mindsets for an Analytically Innovative Organisation," in *International Conference on Decision Support System Technology*, 2018, pp. 103-118: Springer.
[21] T. Rienzo and K. Chen, "Planning for Low End Analytics Disruptions in Business School Curricula," *Decision Sciences Journal of Innovative Education*, vol. 16, no. 1, pp. 23-41, 2018.
[22] Z. Radovilsky, V. Hegde, A. Acharya, and U. Uma, "Skills Requirements of Business Data Analytics and Data Science Jobs: A Comparative Analysis," *Journal of Supply Chain and Operations Management*, vol. 16, no. 1, p. 82, 2018.
[23] P. Mikalef, R. Van de Wetering, and J. Krogstie, "Big Data enabled organizational transformation: The effect of inertia in adoption and diffusion," in *Business Information Systems (BIS)*, Berlin, Germany, 2018.
[24] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big Data*, vol. 1, no. 1, pp. 51-59, 2013.
[25] C. Costa and M. Y. Santos, "The data scientist profile and its representativeness in the European e-Competence framework and the skills framework for the information age," *International Journal of Information Management*, vol. 37, no. 6, pp. 726-734, 2017.
[26] SAS, "Big Data Analytics: An assessment of demand for labour and skills, 2012–2017," SAS2013.
[27] M. D. Myers and M. Newman, "The qualitative interview in IS research: Examining the craft," *Information and organization*, vol. 17, no. 1, pp. 2-26, 2007.
[28] M.-C. Boudreau, D. Gefen, and D. W. Straub, "Validation in information systems research: a state-of-the-art assessment," *MIS quarterly*, pp. 1-16, 2001.
[29] P. Mikalef, M. Boura, G. Lekakos, and J. Krogstie, "Complementarities Between Information Governance and Big Data Analytics Capabilities on Innovation," presented at the European Conference on Information Systems (ECIS), Portsmouth, UK, 2018.