Issues in Information Systems
Volume 20, Issue 4, pp. 65-74, 2019

UPSKILLING AND RETRAINING IN DATA ANALYTICS: A SKILL-ADJACENCY ANALYSIS FOR CAREER PATHS

Ashraf Shirani, San Jose State University, ashraf.shirani@sjsu.edu

ABSTRACT

The ongoing and accelerating pace of digitization and automation, and the resulting demand for knowledge and skills in the associated technologies, have prompted organizations to retrain and upskill their workforce in addition to hiring fresh talent with the requisite skills. Lack of specific career paths for effectively training and upskilling is, however, an obstacle for many organizations, their employees, graduates, and potential candidates to retrain or upgrade their skills . This study examines technical skills required by the leading business and industry employers in the US for the various data analytics fields. Specifically, 50 recent vacancy announcements in five data analytics fields were analyzed for the similarity or adjacency of their skill requirements in order to discover retraining opportunities for employees and graduates to move from one such field to another. Results of data analysis and discussion of the results highlight potential career paths for such transitions.

Keywords: Data Analytics; Data Science; Skill-Adjacency; Up-Skilling; Retraining

INTRODUCTION

With the increasing production and use of data in business and industry, demand for personnel trained in various data and analytics disciplines has been increasing in recent years. Data and analytics are expected to be "the next big wave for talent" (Diplock et al., 2018). IBM estimates the demand for data scientists, data developers, and data engineers to reach about 700,000 positions in the US by the year 2020 (Columbus, 2017).

The digital revolution and accelerating automation of business processes across all industries and sectors of the economy are expected to have enormous implications for training and retraining of the workforce. While creating demand for new skills and competencies, these developments are also gradually rendering the current workforce inadequately prepared for the new roles. A recent report suggests that two-thirds of the IT decision makers believe that there is a gap between their teams' skills and the knowledge required to meet organizational objectives (Global Knowledge, 2018). By 2030, automation and digitization would necessitate change in job categories and roles for as many as 375 million workers worldwide, according to another report (Manyika et al., 2017). Every company today is practically a technology company, though the extent to which the digital technology has permeated their products, services, and business processes varies from one industry to another (Mims, 2018). The inroads of digitization into organizations in certain sectors of the economy are so rapid and accelerating that it is a matter of mere survival for many organizations and their employees to upgrade their infrastructure and skills in short order.

According to a recent survey of 283 private organizations with over $100 million annual revenue by the McKinsey Global Institute, about two-third of the executives in the US and Europe believe that due to digitization and automation, they would need to replace or retrain more than one-fourth of their workforce by 2023 (Illanes et al., 2018). About the same percentage of executives also believes that corporations themselves should lead the initiatives for closing the impending skill gap. Additionally, an overwhelming majority of the executives (82%) believe that at least half of the skill gap should be addressed through retaining and upskilling.

Through the mid 1970s, almost 90% of the US companies filled their vacancies internally, whereas today only less than 1/3rd of them do so (Cappelli, 2019). In order to effectively participate in the digital economy, most companies upgrade their technical infrastructure and hire appropriately trained workers. Training, retraining, and upskilling their existing workforce is often accorded a lower priority by many organizations. This is partly due to the fact that preparing existing workforce for the newer roles is not as expeditious as hiring new workers who already have those skills, and the outcome of the training and upskilling efforts can be uncertain. Researchers and practitioners in many fields,

however, have recently argued for a balanced approach to replenishing organizational talent pools in which both new hires and retrained workers would both contribute to the organizational success through their unique capabilities (e.g., Fraher & Ricketts, 2016; Illanes et al., 2018). Besides satisfying part of an organization's needs for newer skills, retraining and upskilling existing employees confers societal benefits by keeping workers as gainfully employed; it also benefits organizations by retaining the organizational knowledge that their current workforce possesses. Moreover, retraining and upskilling allows organizations to tailor their training programs to meet their specific needs.

Responding to the increasing demand for the newer skills, some organizations are starting to refocus their efforts on retraining and upskilling their workforce (Weber, 2019). Amazon, for example, employs millions of workers in its fulfillment centers. These are typically low-paying and lower skill jobs. Recently, the company has offered some its warehouse workers to retrain them as data technicians and other better-paying positions (Dapena, 2019). After completing a 16-week training and certification program offered at an Amazon's facility, warehouse workers may be hired at the company's data centers. Another major employer, JPMorgan Chase and Company, has teamed up with MIT's Initiative on the Digital Economy to forecast skill needs for the company and develop training programs for its workforce (Weber, 2019). JPMorgan has already implemented the program in its IT department and plans to test it in other departments. AT&T's *Future Ready* program has helped retrain some 180,000 of its employees in partnership with the University of Notre Dame and the Georgia Institute of Technology (Dapena, 2019). Through the *Future Ready* program, AT&T employees have earned short-term badges, nanodegrees, and master's degrees in fields including computer science and data analytics.

A number of large companies are addressing the lack of requisite skills to meet their organizational goals through acquisitions of technology start-ups and smaller companies, though with mixed results. According to a recent study by Kim (2018), about one-third of acquired workers leave the company within one year, and acquired workers are 15% more likely than new hires to leave within three years. One reason for the high rate of departures by acquired employees is organizational mismatch (Somers, 2019). Startups workforce is typically entrepreneurial, risk-taking, and accustomed to working in more autonomous environment with less bureaucracy than more established firms.

Until recently, there have been fewer opportunities to train employees internally in a timely or effective manner, and there are no clear career paths. Also, since many companies have pared down their internal recruiters and have outsourced many recruitment activities, it is harder for them to accurately determine skill sets required for the available positions and attributes of internal candidates who may be a good fit (Cappelli, 2019). One challenge that organizations face in the skill upgrade process is that they need to document exactly what skills each employee currently possesses, specifying the skills the company needs now or in the near future, and devising training programs to fill the gap between the two (Weber, 2019). An approach to facilitating this process suggested by Burning Glass Technologies, a workforce analytics company, relies on the concept of '*skill adjacency'* (Dapena, 2019). By identifying skill-adjacent career opportunities available for a current position to upgrade one's skills and/or education to the next adjacent level, workers and their employers can weigh their upskilling options.

The objective of this study is to explore skill adjacencies among the data disciplines associated with analytics in order to identify and guide potential career paths for advancement from one data analytics field to another. Specifically, recent announcements for analytics positions are examined for their education and skills requirements. The following data analytics fields are included in the study: business analytics; data science; data analytics; data engineering; and big data. We use the terms *analytics* and *data analytics* interchangeably as umbrella terms for these five analytical fields in data discipline. Due to lack of uniform understanding of the terminology involved, it would be helpful to the readers of this study to know how the relevant terminology is used in the context of this study. This terminology is thus summarized below.

Retraining and Upskilling Terminology

The workforce development literature contains a number of terms that are essentially used interchangeably; these include, reskilling, upskilling, retraining, skill-upgrade, and more (Weber, 2019). We use the two terms –retraining and upskilling –to describe the process and attainment of next higher level of knowledge and skills.

Education: Acquisition of knowledge and skills necessary for competency in a discipline. Education may be acquired through formal programs such as bachelor's and master's degrees granted by universities and colleges in the US or

abroad, or it may be obtained through informal and non-traditional sources. An example of the latter type, that is gradually gaining acceptance in the industry, is online education which typically confers *nanodegrees* and certificates of accomplishment.

Knowledge: Understanding of the concepts and methods pertaining to a discipline, obtained through education, experience, and other means.

Skill: Domain-specific proficiency that is often acquired through education, training, or experience. Technical and soft skills are two categories of skills considered in this study.

Competency: The term competency is adapted for this study from the following definition: "A competency is the capability to apply or use a set of related knowledge, skills, and abilities required to successfully perform critical work functions or tasks in a defined work setting" (Krathwohl et al., n.d.). We use the terms competency and capability to refer to the same construct, i.e., the ability to do something.

Putting the above constructs together, the following diagram in Figure 1shows how the relevant education and training components relate to each other and fit in the overall retraining and upskilling landscape.

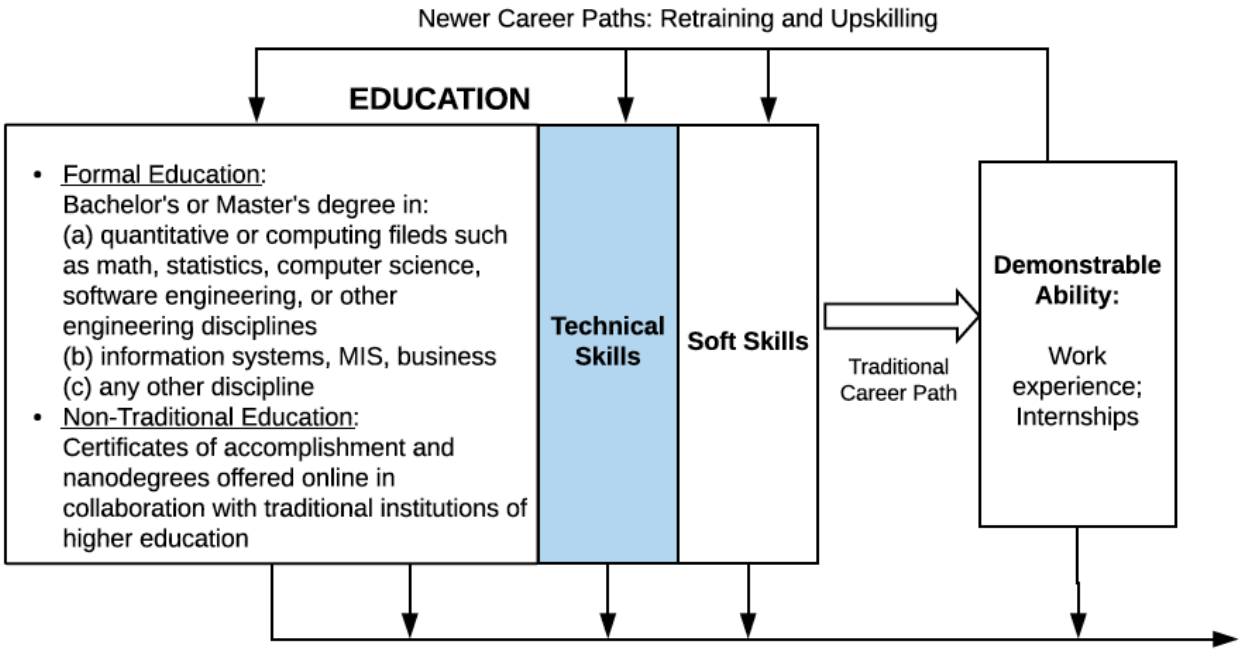

Figure 1: The Competency and Capability Spectrum (The blue shaded area is the focus of this study)

As Figure 1 suggests, the traditional career path to competency and capability begins with education that includes acquisition of knowledge along with technical and soft skills. Work experience adds to one's ability by providing opportunities for further development of technical and soft skills. Such traditional career path has typically assumed that education is acquired through formal studies at a college or university. Gradual acceptance of the newer learning opportunities such as online education by some companies, together with emphasis on retraining and upskilling appear to be the future path to life-long learning and development.

Methodology of the study is described in the next section; the section after that contains results of data analysis and discussion of the results; the last section concludes the paper with career path recommendations, limitations of this study, and suggestions for future research.

METHODOLOGY

Job listings posted on professional analytics websites are expected to reflect skillsets more accurately than those from general jobs portals or private recruiters who are known to post jobs with unrealistic requirements and "phantom" jobs –the jobs that don't actually exist (Cappelli, 2019). Hence, for this study, a total of 50 position announcements were retrieved from online job postings by the following two professional data analytics organizations: *KDnuggets* (https://www.kdnuggets.com); and *Data Science Central* (https://www.datasciencecentral.com). Ten vacancy announcements from each the following five data analytics fields during April, 2019 were retrieved: business analytics; data science; data analytics; data engineering; and big data. Required knowledge and skills for each of the 50 announcements were copied and pasted into text documents, separately for each of the five data analytics fields, as shown in table 1 below. It is useful to indicate the document number here since the data and text mining software used to analyze these documents refers to document numbers in the cluster output.

Table 1. Documents Processed

Document#	Analytics Field
1	Business, business intelligence (BI), and data analyst
2	Big data
3	Data analytics
4	Data engineering
5	Data science

Analysis Stage 1
Using the five text documents, the first stage of the analysis consisted of using text mining software to extract relevant terms and their frequencies. In order to extract only the relevant terms, the following operators were applied during the text mining process:

Tokenization: To extract and count unique words

Removing "stop" words: Words such as conjunctions and articles of speech, and other common words that are not relevant to this study were removed using this sub-process.

Case insensitivity: The text was converted to all lower-case letters to avoid misinterpretation of the same words spelled in different cases.

N-grams: This operator was used to consider two words (such as Apache and Spark) together as a single term rather than two separate ones.

Analysis Stage 2
After generating the five word-count vectors, one each for the five data analytics fields, k-means cluster analysis was performed to explore similarity among the five fields. The value of *k* (the number of clusters) was first set to two clusters, and then to three and four. The purpose of performing three different iterations of cluster analysis was to explore and potentially address the following types of questions through similarity and skill-adjacency analysis:

> *Are data analytics and data science two distinctly different fields in terms of their academic qualifications and technical skill requirements? How are their requirements similar and different from each other?*

Such are some of the questions frequently asked by students, graduates, and practitioners, though definitive answers are not yet available. Some representative answers, though, may be that while data science is an umbrella term for a group of fields including domain expertise, data analytics is a subset of it. Also, data science provides broader insights rather than generating actionable insights (Liberty, 2019). We hoped to provide some clarity and guidance for upskilling through this study.

Likewise, similar questions often arise about *data engineering* and *big data*, or *data analysis* and *data analytics*. Through text mining and cluster analysis, this study aimed to clarify some of such ambiguities by discovering commonalities and differences in academic qualifications and skill requirements among these disciplines.

RESULTS AND DISCUSSION OF THE RESULTS

The text analysis process resulted in a list of terms by each document that were then sorted by their counts. Since the number of ads for each of the five analytical fields was normalized by including equal number of position announcements in each category, ranking of terms generated by the analysis could generally be validly compared among the five fields.

The counts of important terms in each document were then subjected to three different iterations of *k-means* cluster analysis, by setting the value of *k* (i.e., number of clusters) to two, three, and then four. Figure 2, below, compares the output of the three iterations.

2-Cluster Iteration	3-Cluster Iteration	4-Cluster Iteration
root cluster_0 2.0 4.0 5.0 cluster_1 1.0 3.0	root cluster_0 1.0 cluster_1 3.0 5.0 cluster_2 2.0 4.0	root cluster_0 4.0 cluster_1 1.0 cluster_2 3.0 5.0 cluster_3 2.0

Figure 2. Documents Clustered in Three Different Iterations

Below are the significant findings and their interpretations with respect to the objective of this study.

Business Analyst, Business Intelligence, and Data Analyst vs. Data Analytics

The information systems and management information systems (MIS) programs, typically offered by the Colleges and Schools of Business in the US, often prepare their graduates for the first three of these positions, including business analysts, BI analysts, or data analysts. The term analysis used in these designations refers to the process of breaking down the complex systems requirements to simpler parts for further study, design, and development of an information system or application to meet end-user requirements (e.g., Valchanov, 2018). The data analytics filed, on the other hand, focuses on the data as a means to support organizational decisions. Data analytics deals with various aspects of data including collecting, cleansing, and processing to generate valuable insights from it (Chiang et al., 2018).

The three-cluster analysis of this study clearly separates the business/BI/data analysis field from the rest of the fields as it ends up as a unique cluster of its own (Table 2, below):

Table 2. Frequently Required Education and Skills (Business Analytics; BI Analyst; Data Analyst)

Education: Bachelor's or master's in business, MIS, or related field
Technical Skills:
Data visualization; Tableau; Qlik
Data modeling; relational databases; SQL
Microsoft Office, Excel, Excel Solver, Power BI
Project management; CRM; ERP systems; Salesforce
Soft/People Skills:
Influencing; listening; motivation; negotiation; teamwork; problem-solving; customer focus

Data Analytics and Data Science

Both three- and four-cluster iterations of the cluster analysis (Figure 2 above) confirm close association between the two fields in terms of their shared skill requirements. This also indicates that there is some ambiguity, not only on the part of potential candidates but also possibly among the recruiters and employers as to how exactly should the two fields be delineated. Thus, some recruiters possibly prefer to cast a "wider net" in the hope of getting more and better qualified candidates to better their chances of finding suitable fit (Cappelli, 2019).

Table 3: Frequently Required Education and Skills (Data Analytics; Data Science)

Education: Typically requires a bachelor's or master's degree in a quantitative discipline such a Statistics, Math, Data analytics, or Data science; doctorate preferred; occasionally, other degrees are acceptable as well
Technical Skills:
Statistical analyses; predictive modeling; testing; causality research
Python Pandas, NumPy, and other packages; Apache Spark; Hadoop; Machine learning; Scala; Visualization; MATLAB; Relational databases; Deep learning; Optimization; Dimensional modeling;
Soft/People Skills:
communication; global perspective; research; teamwork; problem-solving; collaboration

Data Engineering and Big Data Analytics

The three- and four cluster iterations of cluster analysis (Figure 2 above) are most relevant to understanding similarity and differences between these two fields. The 3-cluster analysis groups these two in a single cluster, indicating close similarity in their job requirements. The 4-cluster analysis, on the other hand, puts each of these in their own separate clusters indicating that that they also have unique requirements. Thus, a high degree of similarity as well as uniqueness would also mean there is a good chance that data engineers, who typically are software engineers as well, would be good candidates for co-skilling in big data technologies. Table 4 summarizes their shared skill requirements and preferred academic qualifications.

Table 4. Frequently Required Education and Skills (Data Engineering; Big Data Technologies)

Education: Typically requires a bachelor's or master's degree in Computer Science, Computer Engineering; or other engineering fields. Occasionally, a degree in information systems or a related field is acceptable.
Technical Skills:
Apache Hadoop, Spark, Hive, Sqoob, NiFi; Java; Scala; HBase; PySpark; Flume; Impala; Parquet; Oozie; Storm; Avro; Relational databases
Less Common: Bitbucket; Columnar and NoSQL databases; Pig; Yarn; Docker
Soft/People Skills:
Written and oral communication; teamwork; problem-solving

Operationalizing the Training and Upskilling Process

Although data analytics skills are the focus of this paper, virtually every organization and most current skillsets will undergo transformation due to increasing automation and digitization. This trend is expected to accelerate over time and is now a significant source of concern among workers and general population (PwC, 2018). It is incumbent upon governments and organizational decision makers to take necessary steps to prepare their workforce for the current and anticipated future demand for skills and competencies. In order to realize this goal, organizations need to undertake a multi-faceted approach that essentially calls for close and continuing collaboration among an organization's decision makers, the human resource department, and functional areas of business. Such a process would involve translating the organization's goals to the skills and competencies the organization needs to achieve those goals; inventorying employees' skills; and developing a strategy for training and upskilling, along with necessary resources and incentives for the employees (Figure 3).

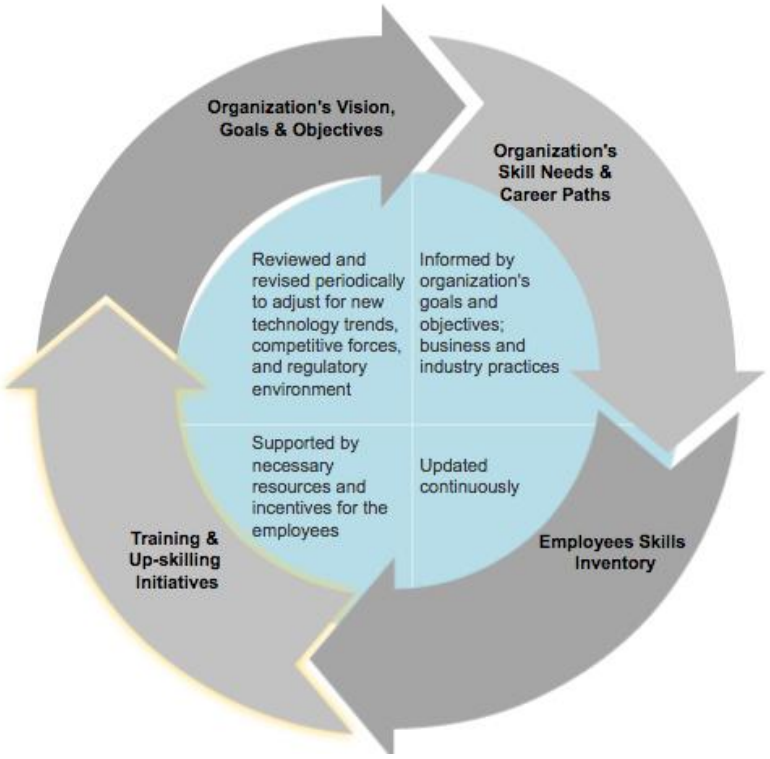

Figure 3. Skills Upgrade Process

The sub-processes in the organization's skills upgrade process are not static or a one-time effort; they are meant to be dynamic and carried out in parallel and continuously.

Regardless of the type of organization, retraining and upskilling often requires a holistic approach involving most, if not all, functional areas of business and workforce at all levels in the organization (Lorica, 2019). For example, top-level executives and business owners need to understand if and where the newer technologies can make an impact in their lines of business; technical personnel need to learn new tools and architectures; business and domain experts need to redesign their business processes and workflows to integrate new technologies; and security and privacy experts need sufficient understanding of the potential concerns and threats that newer systems may pose.

Creating an inventory of employees' skills and keeping it updated is a major step in the skills upgrade process. Despite the increasing role of digitization and automation in business processes, products, and services, many companies don't have a system to track and know what skills their employees currently have or skills they lack to effectively utilize digital technologies (Hess, 2018). There are, however, software tools now available to help. One example is an app

called *Digital Fitness*, by *PricewaterhouseCoopers* (PwC), the second largest professional services firm worldwide. With this app, "employees are able to assess their competence in topics like artificial intelligence, data and analytics, automation and more, as well as engage with relevant learning assets — articles, video content, podcasts and online courses — for areas they need to develop further" (Moore, 2018).

CONCLUSION

The objective of this study was to explore skill adjacencies among the data disciplines associated with analytics in order to identify and guide potential career paths for advancement from one data analytics field to another. A survey of recent vacancy announcements in the data analytics discipline was conducted to gather the latest qualifications and skill requirements for the five constituent fields including business/BI analyst, data analytics, data science, data engineering, and big data. Through text mining of the collected requirements in text documents, and k-means cluster analysis, similarities and dissimilarities among these fields in terms of their requirements were discovered. The analysis reveals the following potential career paths for graduates or employees for career advancement to the next adjacent analytics field.

The upper portion of Figure 4 suggests that information systems, MIS, and business majors may like to opt for career move towards data analytics since academic qualifications pose no obstacle, and skill sets are somewhat similar. Yet doing so would require for such candidates to acquire additional knowledge and skills in quantitative fields through online or traditional academic coursework. A similar move is also possible by data analytics professionals through additional coursework and experience in the domain area of their expertise. For further career enhancement for both of these groups, big data analytics offers a viable path.

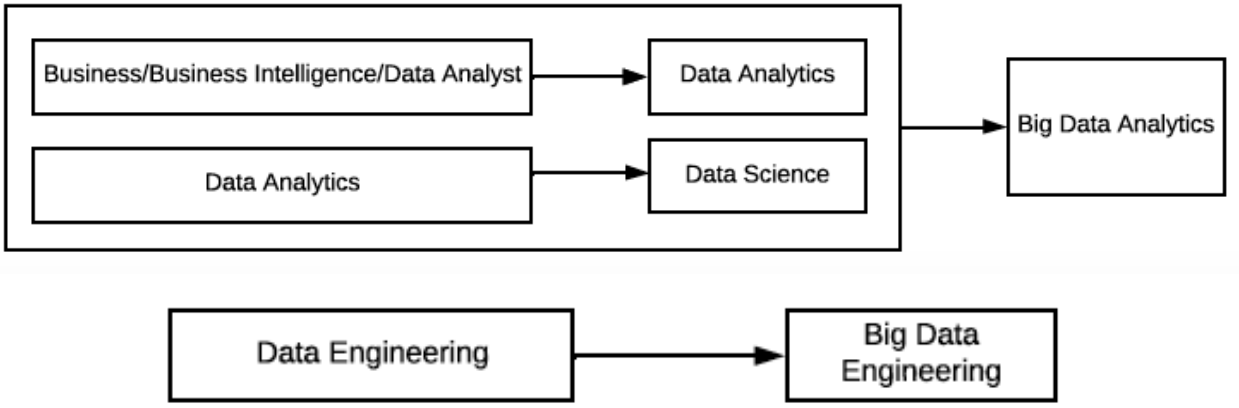

Figure 4. Potential Career Paths

The lower portion of Figure 3 suggests a logical career path for data engineers since both fields have similar degree requirements. Additional coursework and/or experience working with big data technologies offers them a rather quick career transition path.

Limitations and Future Research

This study does have certain limitations that should be taken into consideration. First, the sample size for the study (50 ads) was somewhat limited and expanding it to a larger and more varied sample of postings from additional professional sites would help improve the results in terms of more specificity and granularity of skill development recommendations. Secondly, the terms extracted from the text document were either one or two words each. There appears to be an opportunity for extracting more useful terminology if three-word *n-grams* are used for tokenization in the text mining process.

REFERENCES

Cappelli, P. (2019). Your approach to hiring is all wrong. *Harvard Business Review,* (May-June)

Chiang, R., Grover, V., Liang, T., & Zhang, D. (2018). Special issue: Strategic value of big data and business analytics. *Journal of Management Information Systems, 35*(2), 383-383-387.

Columbus, L. (2017). IBM predicts demand for data scientists will soar 28% by 2020. *Forbes,*

Dapena, K. (2019, 04/20). *The Wall Street Journal,* pp. B6-B7.

Diplock, T., Meier, P., Jordan, D., & Ek, N. (2018). How to actually put your data analysis to good use. *Harvard Business Review,* , May 1, 2019.

Fraher, E., & Ricketts III, T. (2016). Building a value-based workforce in north carolina. *North Carolina Medical Journal, 77*(2), 94-98.

Global Knowldge. (2018). *The future is now with AWS certifications.* https://www.globalknowledge.com/us-en/resources/resource-library/articles/the-future-is-now-with-aws-certifications/

Hess, J. (2018). *Why companies need to build a skills inventory.* https://blogs.oracle.com/oraclehcm/why-companies-need-to-build-a-skills-inventory

Illanes, P., Lund, S., Mourshed, M., Rutherford, S. & Tyreman, M. (2018). *Retraining and reskilling workers in the age of automation.* https://www.mckinsey.com/global-themes/future-of-organizations-and-work/retraining-and-reskilling-workers-in-the-age-of-automation

Kim, J. D. (2019). *Predictable exodus: Startup acquisitions and employee departures.* https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3309324_code2185799.pdf

Krathwohl, D. R., Bloom, B. S. & Masia, B. B. (2010). *Taxonomy of educational objectives.* https://sph.uth.edu/content/uploads/2012/01/Competencies-and-Learning-Objectives.pdf

Liberty, D. (2019). *Data science vs. data analytics – What's the difference?* https://www.sisense.com/blog/data-science-vs-data-analytics/

Lorica, B. *AI and machine learning will require retraining your entire organization* .https://www.oreilly.com/ideas/ai-and-machine-learning-will-require-retraining-your-entire-organization

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., et al. (2017). *Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages*McKinsey Global Institute.

Mims, C. (2018, 12/04/2018). Every company is now a tech company: It's why all established businesses need to hire a technical co-founder. *The Wall Street Journal,* pp. R5.

Moore, E. (2018). *The surprising benefit PwC uses to attract and retain top talent.* https://www.glassdoor.com/employers/blog/pwc-digital-fitness/

PwC. (2018). *Workforce of the future: The competing forces shaping 2030.* https://www.pwc.com/gx/en/services/people-organisation/workforce-of-the-future/workforce-of-the-future-the-competing-forces-shaping-2030-pwc.pdf

Somers, M. (2019). *Your acquired hires are leaving. Here's why.* https://mitsloan.mit.edu/ideas-made-to-matter/your-acquired-hires-are-leaving-heres-why

Valchanov, I. (2018). *Data science vs machine learning vs data analytics vs business analytics.* https://www.kdnuggets.com/2018/05/data-science-machine-learni

Weber, L. (2019, April 20-21, 2019). Evolving at work. *The Wall Street Journal,* pp. B1-B6.