Analytics Prevalent Undergraduate IT Program

Jai W. Kang†
jai.kang@rit.edu

Qi Yu†
qi.yu@rit.edu

Edward P. Holden†
edward.holden@rit.edu

Erik Golen†
efgics@rit.edu

Michael McQuaid†
mjmics@rit.edu

†Rochester Institute of Technology
School of Information
152 Lomb Memorial Drive
Rochester, NY 14623

ABSTRACT

As we enter age of information, computational resources have become relatively inexpensive and seemingly limitless in their capabilities. Consequently, the field of data science has emerged as one of the key areas in industry among engineers, scientists, and IT professionals alike. In this paper, we derive a Data Analytics Body of Knowledge (DA-BoK) from the existing Data Science BoK (DS-BoK) as a means to provide data analytics content throughout the curriculum of an institution's undergraduate IT degree programs. A series of four Knowledge Areas are subdivided into Knowledge Units that can be introduced into existing IT courses through four embedding types, including lectures by example, labs, case studies, and projects. A case study is presented using our three IT programs as examples of how this content can be introduced at three levels of Bloom's Taxonomy (i.e. vocabulary, comprehension, application) through each of the embedding types. Finally, insights and guidance are provided on how to improve the proposed data analytics embedded IT curriculum in order to meet the demands of the modern data analytics pipeline in industry.

CCS Concepts: • Computing education • Computing education programs • Information technology education

Keywords: Information technology; course content; curriculum; teaching methodology; data science; data analytics.

ACM Reference format:
Jai W. Kang, Qi Yu, Edward P. Holden, Erik Golen, and Michael McQuaid. 2020. Analytics Prevalent Undergraduate IT Program. In *Proceedings of ACM Special Interest Group in IT Education Conference (SIGITE'20), October 7-9, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 6 pages.
https://doi.org/10.1145/3368308.3415370

1. Introduction

In searching for new areas to advance the Information Technology (IT) curriculum, IT educators have been exposed to fields such as data science, data analytics, business analytics, data engineering along with big data. Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insight from structured and unstructured data [3]. [2] compared Computer Science (CS) to Data Science (DS) through a discussion of their similarities, differences and overlap. Through a similar approach, Data Analytics (DA), Business Analytics (BA) and Data Engineering (DE) can be compared to IT, Information Systems (IS), and Soft Engineering (SE), respectively. In other words, a pairwise comparison could be made for each new field: CS vs. DS; IT vs. DA; IS vs. BA; SE vs. DE. The World Economic Forum [11] highlights key daily statistics in their infographic using big data: 500 million tweets; 294 billion emails sent; 4 petabytes created on Facebook; 4 terabytes created from each connected car; 65 billion messages sent on WhatsApp; 5 billion searches made; 463 exabytes of data created globally by 2025, which is equivalent to 212,765,957 DVDs per day! This Big Data further highlights and motivates each of these new academic fields: DS, DA, BA and DE.

It could be challenging for IT graduates to exercise their domain knowledge to select, develop, apply, integrate, and administer the big data projects to accomplish their personal, organizational, and societal goals without some level of data analytics knowledge [5]. For example, traditional data management and analysis systems, such as Relational Database Management Systems (RDBMS), are not suitable and adequate to process big data due to the large volume and their unstructured formats [8]. After completing a Data Warehousing course successfully from an IT program, students can develop data warehouses applying the Extract, Transform and Load (ETL) baseline technologies only when all source data are properly structured. It will be challenging however, to create a data warehouse when the source data are in their unstructured raw formats, like customer product reviews, tweets and emails in a product manufacturing and sales environment.

The IT Curricula 2017 (IT2017) [5] recognized the following new emerging disciplines in IT baccalaureate degree programs as supplemental domains: ITS-CEC Cybersecurity Emerging Challenges; ITS-DSA Data Scalability and Analytics. The IT2017 task group recommended that a supplemental domain should account for approximately 4% of the IT curriculum, which could provide students with sufficient exposure to data analytics. However, since not all IT students take the supplemental courses,

ACM ISBN 978-1-4503-7045-5/20/10…$15.00
https://doi.org/10.1145/3368308.3415379

we propose to embed appropriate levels of data analytics in most, if not all IT courses.

The remainder of the paper is organized in 5 sections. We begin by defining Data Analytics Body of Knowledge (BoK) and its derivation from Data Science BoK in Section 2. Section 3 uses a case study to demonstrate how data analytics topics are covered in IT courses irrespective of students' undergraduate IT degree program, such as Computing & IT (CIT), Human Centered Computing (HCC), and Web & Mobile Computing (WMC). Section 4 discusses how the modern data analytics pipeline is integrated for better analysis of large-scale data and offers key insights and important guidance on how to best incorporate analytics topics into existing IT curricula. Section 5 presents the concluding remarks.

2. Data Analytics Body of Knowledge (DA-BoK) (see Tables 1, 2 & 3)

Data Analytics Body of Knowledge (DA-BoK) has been adopted from the Data Science Body of Knowledge (DS-BoK) [12]. The DA-BoK includes four Knowledge Areas (KA), and each KA consists of a number of Knowledge Units (KU) as shown in Table 1. Each KU was mapped to authors' institution IT courses by characterizing how each KU is taught along with students' level of competence after graduation. (Note: authors' institution undergraduate IT courses in three different degree programs will be referred to as '**IT courses'** for the rest of the paper unless otherwise specified. The three undergraduate IT degree programs are CIT (Computing & Information Technologies), HCC (Human Centered Computing) and WMC (Web & Mobile Computing)).

2.1 Body of Knowledge (BoK) (see Table 1)

The Data Science Body of Knowledge [12] consists of five Knowledge Groups including the Data Analytics Group, which is adopted as the Data Analytics Body of Knowledge (DA-BoK) in this study. This DA-BoK suggests 6 Knowledge Areas (KA): Statistical methods; Machine Learning; Data Mining; Text Mining; Predictive Analytics; and Computational Modeling, Simulation & Optimization. Since Data Mining is applications of Machine Learning, Machine Learning is removed in this study. Then Knowledge Units (KU) in each KA are mapped to all IT courses for potential applications of the KU in the course. When no mapping is found between KU and IT course, the KU is removed from the study.

2.2 Competence

The lecture coverage level of each Knowledge Unit (KU) was classified using Bloom's taxonomy [1], which represents three levels of capability of a student after graduation [6]:

1. 'v' (vocabulary): a student should be able to understand the terminology to delegate to others if necessary;
2. 'c' (comprehension): a student should be able to intelligently discuss the topic and perform basic tasks;
3. 'a' (application): a student should be able to apply knowledge to perform tasks competently as required in a work environment.

Table 2 shows Bloom's taxonomy level of analytics coverages for all IT courses.

2.3 Embedding Types

As [7] discusses the case of information security curricula, five approaches can be considered when including Data Analytics in the IT Programs in the iSchool: 1) add knowledge units (KU) to existing courses called *embedding*, 2) add KU's to capstone courses, 3) create independent Data Analytics courses, 4) create Data Analytics certificates/minors, and 5) create a Data Analytics degree program. In this study we propose mechanisms to apply the *embedding* approach. Table 3 shows 4 different ways of embedding each KU into the IT courses: 1) Lecture by Example (LBE); 2) Lab; 3) Case Study; 4) Project. KU may also be embedded into a course in multiple ways, depending on the particular combination of KU and IT course content.

Table 1. Data Analytics Body of Knowledge

Knowledge Area (KA)		Knowledge Units (KU)	
1	**Data Mining**	1.1	Supervised Machine Learning
		1.2	Unsupervised Machine Learning
		1.3	Classification methods
		1.4	Data mining and knowledge discovery
		1.5	Knowledge Representation and Reasoning
		1.6	CRISP-DM and data mining stages
		1.7	Anomaly Detection
		1.8	Time series analysis
		1.9	Feature selection, Apriori algorithm
		1.10	Graph data analytics
		1.11	Visual Analytics; Dashboard
2	**Text Data Mining**	2.1	Text analytics including statistical, linguistic, and structural techniques to analyze structured and unstructured data
		2.2	Data mining and text analytics
		2.3	Natural Language Processing
		2.4	Sentiments analysis
3	**Big Data Infrastructure & Tech**	3.1	Computer systems organization for Big Data applications, CAP, BASE and ACID theorems
		3.2	Parallel and Distributed Computer Architecture
		3.3	High Performance and Cloud Computing
		3.4	Clouds and scalable computing
		3.5	Cloud based Big Data platforms and services
		3.6	Big Data (large scale) storage and filesystems (HDFS, Ceph, etc)
		3.7	NoSQL databases
		3.8	Computer networks: architectures and protocols
		3.9	Big Data Infrastructure management and operation
4	**Project Management**	4.1	Project Management

3. Case Study

There is increasing demand for BS graduates with knowledge of data analytics in all disciplines. Many existing IT courses, such as database management, contain material that can form the foundation that data science modules can build upon to get more deeply into data analytics processes. In this case study, we will examine our IT programs to determine where embedding data analytics topics in existing courses can benefit our students in our three BS programs: CIT, HCC and WMC. We will tie this embedded data analytics content to the knowledge areas (KA) and the knowledge units (KU) specified in Table 1.

Table 2. Data Analytics Embedded IT Courses

					Data Analytics Body of Knowledge (DA-BoK)																								
IT Deg				**Competence**	**Data Mining (KA#1) KU's**											**Text Mining (KA#2) KU's**				**Big Data (KA#3) KU's**									**Proj Mgt (KA#4)**
BS	**Course #**	**Core**	**Course Title**	**Bloom's Taxonomy**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**1**
CIT	ISTE-121	*	Computational Problem Solving in the Information Domain II	v																									x
	ISTE-230	*	Introduction to Database and Data Modeling	v						x																			
	ISTE-430	*	Information Requirements Modeling	c																									x
	ISTE-434		Data Warehousing	v												x	x	x	x						x	x			
	ISTE-438		Contemporary Databases	c					x					x						x	x	x				x			
	ISTE-470		Data Mining and Exploration	a			x				x		x																
	NSSA-220	*	Task Automation Using Interpretive Languages	v				x			x	x			x														
	NSSA-241	*	Introduction to Routing and Switching	v								x																	
	NSSA-245		Network Services	v				x			x				x														
	NSSA-320		Configuration Management	v																								x	
	NSSA-322		Systems Administration II	c/a																			x	x	x				
	NSSA-370		Project Management	c																									x
	NSSA-425		Data Center Operations	v				x	x	x			x															x	
	NSSA-445		Foundations of IoT	v	x	x		x			x	x			x														
HCC	ISTE-121	*	Computational Problem Solving in the Information Domain II	v																									x
	ISTE-262	*	Foundations of Human Centered Computing	v			x								x														
	ISTE-264	*	Prototyping and Usability Testing	v			x								x														
	ISTE-266	*	Design for Accessibility	v			x								x														
WMC	ISTE-121	*	Computational Problem Solving in the Information Domain II	v																									x
	ISTE-230	*	Introduction to Database and Data Modeling	v						x																			
	ISTE-358		Foundations of Wearable and Ubiquitous Computing	v				x			x	x			x														
	ISTE-430		Information Requirements Modeling	c																									x
	ISTE-438		Contemporary Databases	c					x					x						x	x	x				x			
	ISTE-444		Web Server Development and Administration	v											x														
	ISTE-458		Advanced Topics in Wearable and Ubiquitous Computing	v				x			x	x			x														
	ISTE-470		Data Mining and Exploration	a			x																						

3.1 Computing and Information Technologies (CIT)

The BS in Computing and Information Technologies (CIT) is designed to expose students to a wide variety of computing topics, including web design, databases, computer networking, and system administration, to allow students to become a jack of all trades. Students then choose from one or more concentrations in the areas of database applications, enterprise administration, networking and communications, and web development to hone their skills.

Table 3. Analytics Embedding Types

		Embedding Types			
BS	Course #	Lec By Example (LBE)	Lab	Case Study	Project
CIT	ISTE-121	x			
	ISTE-230	x			
	ISTE-430	x		x	
	ISTE-434	x	x	x	
	ISTE-438	x	x		
	ISTE-470	x	x	x	x
	NSSA-220	x			
	NSSA-241	x			
	NSSA-245	x			
	NSSA-320		x	x	
	NSSA-322		x		
	NSSA-370	x		x	x
	NSSA-425	x			
	NSSA-445	x	x		
HCC	ISTE-121	x			
	ISTE-262	x	x		
	ISTE-264	x	x		
	IISTE-266	x	x		
WMC	ISTE-121	x			
	ISTE-230	x			
	ISTE-358	x			
	ISTE-430	x			
	ISTE-438	x	x		x
	ISTE-444	x			
	ISTE-458	x			
	ISTE-470	x	x	x	x

Currently, there are two analytics courses in the database applications concentration, Data Mining and Exploration and Data Warehousing, both of which have shown significant growth in enrollment in recent years as students have become keenly aware of their marketability in landing both cooperative education and fulltime positions.

Beyond these two courses, we believe that there are several opportunities to embed data analytics content both in the core of the CIT program as well as in the enterprise administration and networking and communications concentrations. Within the core, the rationale is to introduce some basic terminology of data analytics at the Vocabulary level of Bloom's Taxonomy in the form of lectures by example from multiple perspectives, primarily from data mining concepts (KA 1). This includes noting the place of data analytics within a project and its impact on planning a timeline in addition to ingraining the notion of utilizing data from databases or other means of data collection for future analysis. The latter is especially important from a data analytics perspective, as students may not understand the value of the data they have stored or collected from a business perspective.

Within the enterprise administration and networking and communications concentrations, embedding opportunities range from additional domain specific terminology to implementation of data analytics infrastructure. In this era of information, professionals in networking, security, and systems administration need to go beyond simply inspecting log files and network traffic by hand or through basic dashboards so that knowledge can be automatically extracted from these artifacts and acted upon. System administrators are also key figures in making it possible for analysts to perform their duties by building sandboxes and other computing environments for the analysts. As such, there is a spectrum of embedding types associated with concepts from Big Data Infrastructure & Technology (KA3) ranging from Vocabulary level lectures by example to Competence level labs and case studies to an Application level project in System Administration II.

3.2 Human Centered Computing (HCC)

Human centered computing (HCC) introduces students to design and research, which prepares them for careers as designers and researchers. This discipline exists at the intersection of front-end development, psychology, and design. Design includes user interface design, interaction design, and user experience (UX) design. Students in this field are exposed to design thinking and the iterative approach to designing, building, and evaluating digital artifacts such as user-facing applications and websites. This is a multi-disciplinary program, with core courses taught in the iSchool, the Department of Psychology in the College of Liberal Arts, and the School of Design in the College of Art and Design.

The need for data analytics appears mainly in the evaluation phase of the HCC student's workflow. While formerly students were taught mostly qualitative methods of design, the ready availability of logs and clickstream data makes it possible to employ quantitative methods as well. The problems are mainly classification problems (KU1.3), such as which design attribute leads to more conversions (achievements of a desired behavior such as click-through or purchase).

While formerly students were well-equipped if they learned to simply conduct chi-squared tests on contingency tables, available data today encourages the use of techniques like logistic regression and support vector machines (SVM) to classify visitors to a website or users of an application (KU1.3). Additionally, data visualization techniques (KU1.11) offer value for both exploring data and reporting results of classification efforts. These techniques benefit from the increasing number of attributes available for study on consumers.

The value of data analytics to the HCC student spans predictive and descriptive analytics. Predictive analytics seeks to both predict future results and to discern patterns in data not evident from simple analytic methods, such as crosstabulation. Descriptive analytics on the other hand, concerns the identification and reporting of patterns and trends in historical data. All HCC core courses taught in the iSchool are in the laboratory environment, so a combination of lecture and lab embedding types suffices to educate students to the (v) level in Bloom's taxonomy.

3.3 Web and Mobile Computing (WMC)

The primary focus of the WMC degree is to produce skilled developers of web and mobile applications. Students learn how to integrate the back end with the front-end user experience, across several languages and platforms, to impact the application design process at all levels. In the Database concentration, students already have the opportunity to take the Data Mining and Exploration course which covers topics in data analytics extensively (KU1.3, 1.7 & 1.9). These topics include introduction to data mining, data attributes, data preprocessing, data exploration and visualization, classification techniques, classification model analysis and evaluation, clustering techniques, cluster analysis, and association rule mining. This course raises them to the application (a) level of competence on Bloom's taxonomy. The database concentration also has Contemporary Databases, which currently covers NoSQL databases at an application (a) level (KU3.7) including the document and graph database models. We could add deeper material on BASE and CAP and introduce a lab on data analytics that would give the student a vocabulary (v) level of understanding (KU1.5& 3.1).

However, there are also a good number of students who do not take this course. It is our desire that all students learn at least the vocabulary (v) of data analytics as indicated in section 2.2. This is where the embedding technique comes in.

Two required courses give us an opportunity to do this. In Introduction to Database and Data Modeling, we can introduce different vocabularies (v) needed for data analytics. We currently spend time on the ACID principles, but other integrity models such as CAP and BASE could also be added. These are used in newer databases common to web and data analytics applications (KU 1.6, 3.1). In the programming courses, currently using Java, we can add exposure to languages more appropriate to data analytics, such as Python and also introducing project management for data analytics (KU4.1).

In the two elective courses in the Wearable and Ubiquitous Development concentration, we could introduce vocabulary (v) level lectures on data mining and knowledge discovery, anomaly detection, time series analysis, or visual analytics (KU1.4, 1.7, 1.8 & 1.11). In the Web Applications Development concentration, the Web Server Development and Administration course, we could add a lecture on configuring and administering a server for processing data for large scale visualization (KU1.11). In Project Lifecycle Concentration, Information Requirements Modeling course, a lecture and a case study could be added to give students comprehension (c) level of understanding of CRISP-DM or other project management methods used in data analytics (KU4.1).

By making these changes to the courses in the program, all students would have a vocabulary level of exposure to Data Mining (KA1), and Project Management (KA4). Depending on the concentration selected, students could have a vocabulary (v) or an application (a) knowledge level in Data Mining (KA1). They could also have a vocabulary (v) or application (a) level in Big Data Infrastructure and Technology (KA3) and Project management (KA4).

4. Discussion and Recommendation

The classical process of knowledge discovery in databases (KDD) consists of three key phases: data preprocessing, data mining/machine learning, and postprocessing. In particular, data preprocessing is to convert raw data into appropriate forms that can be analyzed by various data mining models. Commonly adopted techniques include feature selection/extraction, dimensionality reduction, subsampling, missing value/noise removal, and so on. Once the raw data is processed, they can be fit into different data mining models for pattern mining and knowledge discovery, aiming to gain valuable information from large-scale data through statistical analysis. Commonly used data mining/machine learning models include supervised learning, unsupervised learning, and reinforcement learning. Finally, postprocessing aims to deploy the learned model to a production environment so that the discovered information/knowledge can benefit actual decision-making in real-world settings.

With large volumes of data being collected and the nature of the data becoming much more complicated, the boundaries among these three phases have become much less distinct. For example, data visualization has been primarily used in the postprocessing phase of the classical KDD pipeline to facilitate interpretation of the result and communicate the discovered knowledge. However, visualization has turned into an essential tool for data preprocessing as it provides an effective way to perform exploratory analysis of complex data, such as gene sequences or other scientific datasets. Such exploratory analysis is of significant importance to understand the nature of the data that helps extract important features so that a meaningful analysis can be performed. Another example is feature selection and extraction, which is more commonly performed directly by data mining/machine learning models while making predictions. Indeed, selecting/extracting features without relating to the downstream data analysis tasks may result in less useful or even noisy features that negatively impact the performance of data mining/machine learning models.

Given these facts, we argue that the modern data analytics pipeline is a **more tightly coupled** one, where key data analytics techniques may appear in various stages of the pipeline and different techniques are **more strongly integrated** for better analysis of large-scale data with ever-increasing complexity.

As the different techniques in the modern data analytics pipeline show a stronger interdependency, when teaching these topics to students, it is important not to cover a specific topic in isolation. To this end, embedding these topics into relevant courses offers some key advantages. First, embedding provides the natural context that shows where these techniques will be needed, which will help students understand not only the technique themselves but also their purpose and potential use cases. Second, the connections between the data analytics techniques and other relevant techniques, such as programming, statistics, data management, web development, human computer interaction, are clearly established. These connections will help students better prepared for a world we are quickly moving into, where data analytics will be ubiquitously applied. Third, embedding also offers a flexible way to cover the same topic multiple times and at different levels. This can ensure all the students can be exposed to basic analytics concepts that all future IT professionals should be familiar with. Meanwhile, for students who show strong interest in data analytics, they have the opportunity to further develop a deeper knowledge by taking advanced elective or concurrent courses.

There are a few very interesting observations from Table 2 of the case study. These observations can also offer some guidance on how to better embed analytics topics in an existing IT curriculum. First, an early embedding could be further strengthened in most programs. As can be seen, the analytics concepts are only sparsely covered at the 100-200 level courses, which implies that most students need to wait until their junior year to get exposed to this key area. Second, text data mining, as a whole knowledge area, should play a much stronger role in the IT curriculum given that most of the data in the modern world are in some unstructured format. Text is one of the most important unstructured data, in the forms of web pages, social media, emails, notes, and many more. The ability to model, store, access, and analyze large-scale text has become one of the most essential skills for future IT professionals. Third, analytics should be more comprehensively integrated into the human centered computing program to better support various human activities. The HCC area is largely empty in the current embedding map, which indicates the lack of support in this critical area for training next generation professionals in human computer interaction. Finally, some specific topics of significant importance are not adequately covered in the current programs. For example, time series analysis should carry a much higher weight in modern data analytics as many types of data, such as network traffic, stock exchanges, sensor readings, videos, speeches, among many others, are usually collected in real-time and fast changing. Students should be equipped with the necessary analytics skills to handle dynamic and fast changing data.

In addition to the above observations and guidance, different embedding types in Table 3 as discussed in our case study also suggest proper ways to incorporate analytics topics in an IT curriculum, especially into some existing courses. Among the four embedding types, LBE offers a convenient and light way to embed analytics topics that build upon the existing topics in a course but make extensions and connections to analytics topics. Thus, the change to an existing course will be minimum. For courses that cover general topics that are related to data, labs and case studies, are suitable choices given that analytics has become an essential way to deal with data. Finally, for courses that already focus on analytics, a semester-long project is usually necessary to equip students with sufficient skills to tackle real-world data analytics problems.

5. Conclusion

In this paper, we extend the Body of Knowledge for Data Science to formulate a formal BoK for Data Analytics, in recognizing the fast-increasing market demand in this key area. We propose to embed the key knowledge units from the four major knowledge areas in analytics into existing IT curricula. We further suggest different embedding types to cover these KUs, including LBE, lab, case study, and project, to ensure expected competency for the corresponding analytics skills of a student upon graduation. A case study is then provided to showcase how the relevant analytics KUs can be embedded into an existing IT undergraduate curriculum. Finally, some key insights and guidance are offered to further improve the embedded IT curriculum in order to meet the new requirements raised by the modern analytics pipeline.

REFERENCES

[1] Canada's Association of Information Technology (IT) Professionals, "A Guide to the Common Body of Knowledge for Computing and IT (CBOK)", Canadian Information Processing Society, Canada retrieved May 24, 2019 from http://www.cips.ca/CBOK.

[2] Computer Science vs. Data Science. Retrieved on 6/6/20 from https://www.onlineengineeringprograms.com/faq/computer-vs-data-science

[3] Data Science (n.d.). In Wikipedia. Retrieved June 1, 2020, from https://en.wikipedia.org/wiki/Data_science.

[4] I. B. Hassan and J. Liu, "Embedding Data Science into Computer Science Education," 2019 IEEE International Conference on Electro Information Technology (EIT), Brookings, SD, USA, 2019, pp. 367-372, doi: 10.1109/EIT.2019.8833753.

[5] IT2017 Task Group. 2017. Curriculum Guidelines for Baccalaureate Degree Programs in Information Technology. DOI: http://it2017.acm.org.

[6] Jai W. Kang, Qi Yu, Edward P. Holden and Xumin Liu. 2019. Complementing Course Contents Between IT/CS: A Case Study on Database Courses. In *Proceedings of ACM Special Interest Group in IT Education Conference (SIGITE'19), October 3-5, 2019, Tacoma, Washington, USA.* ACM, New York, NY, *USA, 6 pages.* https://doi.org/10.1145/3349266.3351414.

[7] Jai W. Kang, Qi Yu, Edward P. Holden, and Tae H. Oh. 2016. Security Requirements Embedded in MS Programs in Information Sciences and Technologies. In Proceedings of the 17th Annual Conference on Information Technology Education (SIGITE '16). Association for Computing Machinery, New York, NY, USA, 77–82. DOI:https://doi.org/10.1145/2978192.2978235.

[8] M. Tanwar, R. Duggal and S. K. Khatri, "Unravelling unstructured data: A wealth of information in big data," 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, 2015, pp. 1-6, doi: 10.1109/ICRITO.2015.7359270.

[9] Malcolm Haynes, Joshua Groen, Eric Sturzinger, Danny Zhu, Justin Shafer, and Timothy McGee. 2019. Integrating Data Science into a General Education Information Technology Course: An Approach to Developing Data Savvy Undergraduates. In Proceedings of the 20th Annual SIG Conference on Information Technology Education (SIGITE '19). Association for Computing Machinery, New York, NY, USA, 183–188. DOI:https://doi.org/10.1145/3349266.3351417

[10] Michael E. Whitman and Herbert J. Mattord. 2004. Designing and teaching information security curriculum. In Proceedings of the 1st annual conference on Information security curriculum development (InfoSecCD '04). Association for Computing Machinery, New York, NY, USA, 1–7. DOI:https://doi.org/10.1145/1059524.1059526.

[11] World Economic Forum and Jeff Desjardins. 2019. How much data is generated each day? Retrieved on 6/6/20 from https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

[12] Y. Demchenko, T. Wiktorski, J. Cuadrado Gallego and S. Brewer, "EDISON Data Science Framework (EDSF) Extension to Address Transversal Skills Required by Emerging Industry 4.0 Transformation," 2019 15th International Conference on eScience (eScience), San Diego, CA, USA, 2019, pp. 553-559, doi: 10.1109/eScience.2019.00076.