*International Journal of Research in Engineering and Science (IJRES)*
*ISSN (Online): 2320-9364, ISSN (Print): 2320-9356*
*www.ijres.org Volume 4 Issue 7 ǁ Jul. 2016 ǁ PP.83-91*

# A New area of Computer science in field of emmbeded system engineering

D. Bright Anand, Nirupama Tripathy, Radha Mohan Acharya, Anita Subudhi
*Department of Computer science and Engineering, NM Institute of Engineering and Technology,Bhubaneswar , Odisha*
*Department of Computer science and Engineering, Raajdhani Engineering College,Bhubaneswar,Odisha*
*Department of Computer science and Engineering ,Aryan Institute of Engineering and Technology Bhubnaeswar , Odisha*
*Department of Computer science and Engineering,Capital Engineering College*

***ABSTRACT-*** *As a fast-growing field with widespread impact on economy and its promising hiring potential, Data Science has been enthusiastically sought-out by many disciplines in the academia, including Mathematics, Statistics, Library Science, Management Information Systems, as well as Computer Science, as an attracting area for recruiting students, soliciting grants, and expanding their existing programs. Although some standalone programs in Data Science have been established around the country, many institutions encounter the challenges in balancing the curricula among different disciplines, allocating new resources, and cooperating with the existing majors regarding enrollment management, student advising, and faculty preparation. In this paper, a strategy-based framework is proposed for those who have limited resources of all kinds to introduce Data Science through their existing Computer Science programs at a baccalaureate level with a minimum curriculum disruption. Instead of a standalone program, to embed Data Science into Computer Science education is demonstrated to be a practical, effective, cost-saving approach based on an extensive study of the synergy between Data Science and Computer Science education. While the proposed framework is not a one-size-fits-all approach, it provides a doable route for blending Data Science into Computer Science education in systematic ways. It has become the consensus that an adequate exposure to Data Science will better prepare computer science students for taking the challenges in this ever-changing, data embraced world.*
*K**EYWORDS**: Data Science Education, Computer Science Education, Curriculum Design*

## I. INTRODUCTION

According to a study discussed in [10], approximately 1,200 exabytes of data are generated annually. This resulted the concept of Big Data, which is a huge amount of unstructured and structured data that has the potential to discover new information. The massive amount of unstructured data presents enormous challenges for the business and IT sectors. It requires specialized skill sets to decipher vast magnitudes of data, and one of these skills is the ability to analyze. Data science, business intelligence, analytics, and other related fields in big data have become increasingly vital in both academic and business communities. Big data has been used to transform medical practice, modernize public policies, and inform business decision-making [12].
It is experiencing rapid and unplanned growth, spurred by the proliferation of complex and rich data in science, industry, and government. Fueled in part by reports, such as the widely cited McKinsey report [11], that forecast a need for hundreds of thousands of data science jobs in the next decade [5]. The need for new and innovative tools for managing and deriving insights from big, unstructured data continues to grow. Consequently, there is a rapidly increasing demand for data scientists who know how to apply these new tools to handle big, unstructured data and to solve business problems [2].

As the global demand for data analysts continue to grow, colleges and universities rushed to offer programs to equip their graduates with the necessary analytical skills [15]. In response to this novel demand, several U.S. higher education institutions have launched blended programs such as Data Analytics, Business Intelligence, or Data Science [16].

For a standalone program in Data Science, a full spectrum of degree programs has been established in the country. For instances, a 72-credit Ph.D. program in Data Science offered by Data Science Center at New York University [23], and a 78-credit Ph.D. program in Analytics and Data Science, available at Analytics and Data Science Institute at Kennesaw State University, Atlanta, Georgia [24]. Master programs in Data Science have been extensively studied in [16]. Based on 30 randomly selected MS programs in data science or a related area, such as Data Analytics or Health Informatics, it is concluded that the required courses are varied, except for those covering the basic analytic skills, from disciplines where the programs are offered [16]. As the bachelor programs in Data Science offered by Math and Statistics Departments can be found in [25, 26], while the programs offered by Computer Science Departments can been seen in [27, 28].

Although it might be the best way to promote Data Science education through a standalone program, many are facing the challenges in the constraint in budget, infrastructure, student recruitment, and faculty preparation [9]. As the importance and the critical role of Data Science to our students, it is needed to find a practical way to equip the fundamental knowledge and skills of Data Science to our computer science students. The rest of the paper is arranged as follow. After an overview of Data Science in section 2, the curriculum guidelines and framework is discussed in section 3. In section 4, the synergy between computer science and data science is analyzed, and the recommendation is outlined in section 5. Finally, in section 6, the conclusion and the future work is presented.

## II. AN OVERVIEW OF DATA SCIENCE

Several reports have revealed the demand for data science. Manyika et al. [11] at McKinsey Global Institute (MGI) predict that by 2018 the U.S. could face a serious shortage of deep analytical skills required for as many as 190,000 positions. The U.S. will not be able to fill this gap simply by changing graduation requirements, waiting for students to graduate with more Data Science skills, or by importing the talent from overseas. Therefore, data science and data analytics fields have been developed as to deal with the big data tsunami. The data science concept unifies statistics, data analysis and their related methods in order to understand and analyze actual phenomena with data. Data Science can be defined as being comprised of three areas: analytics, infrastructure, and data curation [7].

The data life cycle and surrounding data ecosystem from the NSF CISE (2016) [18] report presented Data Life Cycle. Figure 1 shows a simplified diagram with essential components of the data life cycle. Data is generated from a source, cleaned and edited, used/re-used analysis, published or disseminated its findings and finally preserved as illustrated in Figure 1. Additionally, there are other factors including data quality, ethics, policy or regulators, platform and infrastructure that might have an impact on any data.

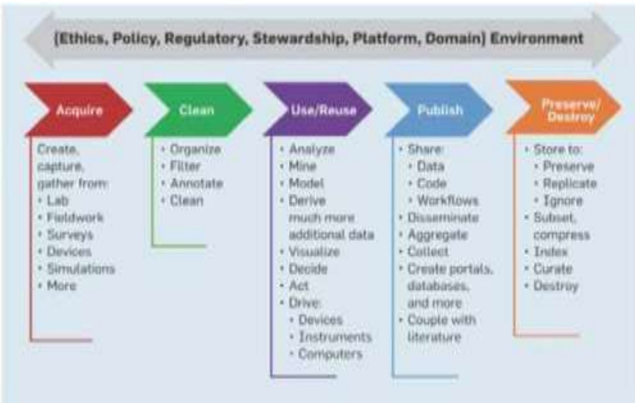


**Figure 1:** The data life cycle and surrounding data ecosystem [18]

The Data Science Competency Framework (CF-DS) was part of the EDISON project that was conducted in 2017 [6]. It focused a foundation for establishing a new profession of data scientist for European research and industry. The EDISON DS framework(EDSF) components consist of Data Science Competence Framework (CF-DS), Data Science Body of Knowledge (DS-BoK) and Data Science Model Curriculum (MC-DS), and Data Science Professional Profiles (DSPP). The EDSF provides a conceptual basis for the Data Science Profession definition targeted education and training, professional certification, organizational and individual skills management and career transferability [6].

The proposed EDSF provides a baseline for developing DS curriculum and course selection with the required competences and skills for a specific industry domain [6]. Figure 2 below illustrates the main components of the EDISON Data Science Framework (EDSF) and their interrelations that provides conceptual basis for the development of the DS profession.

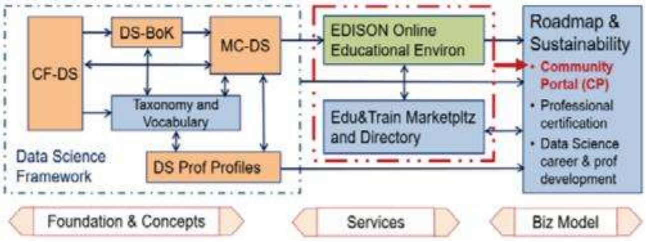


Figure 2: EDISON Data Science Framework components [6]

Domain scientists focus on generating data and using it. Computer scientists often focus on platform issues, including mining, organizing, modeling and visualizing, and the potential for extracting insights from analysis of the data through machine learning (ML) and other approaches [1]. There is also an opportunity to bridge gaps among ML, data analytics, and synergistic partner disciplines such as statistics and operations research. The new opportunity is to focus on the full data life cycle [1].

In addition to the effort made by the academia from the U.S., the faculty members from the European Union drafted their Data Science framework through the EDISON project [6]. Although the proposed framework is based on the structure introduced by NIST [13], the relationship and transition between the competence groups demonstrated in the framework provide a solid foundation for designing the corresponding curricula.

This framework provides basis for other components of the Data Science professional ecosystem such as 1) EDISON Online Education Environment (EOEE); 2) Education and Training Directory and Marketplace; 3) Data Science Community Portal (CP) that also includes tools for individual competences benchmarking and personalized educational path building; 4) Certification Framework for core Data Science competences and professional profiles.

## III. CURRICULUM GUIDELINES AND FRAMEWORKS

The Park City Math Institute (PCMI) 2016 held a three-week workshop for a 25 undergraduate faculties focused on the task of producing curriculum guidelines for an undergraduate degree in Data Science. The group consists of statisticians, mathematicians and computer scientists from a variety of liberal arts colleges and research universities [5]. The group discussed what activities and skills they thought would be necessary for DS program and how they can be implemented currently and in the future. They proposed an undergraduate Data Science guideline on a ten semester-course major common among the liberal arts colleges.

The group highlighted that DS students must encounter frequently project-based, real-world applications with real data to complement the foundational algorithms and models which is a similar view of the Committee for the Undergraduate Program in Mathematics Curriculum Guide 2004-2015 [5]. The group proposed a Bachelor of Science Data Science

(BSDS) program that consists of six competent areas plus a domain related area as shown in Table 1 below.

**Table 1**: Competencies for BSDS Programs [5]

| | Competence | Description |
|---|---|---|
| 1 | Computational Thinking | Data Science graduates should be proficient in many of the foundational software skills and the associated algorithmic, computational problem solving of the discipline of computer science. |
| 2 | Statistical Thinking | It is an approach to understanding the world through data, and involves everything from problem formulation through conclusions (Wild & Pfannkuch) |
| 3 | Mathematical Foundations | The emphasis of an undergraduate Data Science degree should be on choosing, fitting and using Mathematical models. |
| 4 | Model Building and Assessment | Informal Modeling: involves identifying potential sources of variation, discerning between stochastic and deterministic variation, and understanding how these might be modeled mathematically and computationally. Formal Modeling: graduates should be able to build and assess statistical and machine learning models, employ a variety of formal inference procedures, and draw appropriate scope of conclusions from the analysis. |
| 5 | Algorithms and Software Foundation | The Data Science graduate should be able to employ algorithmic problem solving skills to the task. These include defining clear requirements to a problem, decomposing the problem, using efficient strategies to arrive at an algorithmic solution and implementing solutions through programming in a suitable high-level language. |
| 6 | Data Curation | Data curation involves managing data through the entire problem-solving process. |
| 7 | Knowledge Transference–Communication and Responsibility | Data do not exist in a vacuum, but arise from a particular context. Knowledge of that context is necessary to analyze the data, and thus undergraduates need experience applying their discipline outside the core of statistics, computing, and mathematics. |

prerequisites required, integrative, cross-disciplinary collaboration, a grassroots and faculty-driven effort, responsive to students, committed to diversity and inclusion, an open-source approach and an education program inextricably linked to research [4].

The curriculum was built upon three interrelated domains: inferential thinking, computational thinking, and critical engagement with questions of real-world relevance. Table 4 below shows how “Data Analysis,” “Statistics,” and “Computing” are connected by Berkeley’s faculty.

**Table 4:** Modular-and-Connector based course structure [4]

| Data Analysis | Statistics | Computing |
|---|---|---|
| Visualization<br>Sampling<br>Statistic Concepts<br>Hypothesis Testing<br>Prediciton<br>Regression<br>Classification | Probability<br>Permutation<br>Data Distribution<br>Statistic Laws<br>Statistic Properties<br>Statistic Testing<br>Statistic Interpretation | Data structures<br>Data Representation<br>Process Control<br>Databased Construction<br>Database Query<br>Software Development<br>Programming languages |

To match the competencies identified in Table 1, 9 courses are proposed in [5] as shown in Table 2 below:

**Table 2**: Nine Proposed Courses for BSDS program

| | Proposed Courses |
|---|---|
| 1) | Introdution to Data Science I |
| 2) | Introduction to Data Science II |
| 3) | Mathematics for Data Science I |
| 4) | Mathematics for Data Science II |
| 5) | Intro to Statistical Models |
| 6) | Statistical and Machine Learning |
| 7) | Algorithms and Software Concepts |
| 8) | Databases and Data Management |
| 9) | Capstone Course |
| | Course in an Outside Discipline |

The matching between the proposed courses and the BSDS competencies is demonstrated in Table 3 below.

**Table 3:** Course Matching to Competencies

| | Competencies | Courses |
|---|---|---|
| 1 | Computational Thinking | - Algorithms and Software Concepts<br>- Databases and Data Management |
| 2 | Statistical Thinking | - Intro to Statistical Models<br>- Statistical and Machine Learning |
| 3 | Mathematical Foundations | - Mathematics for Data Science I<br>- Mathematics for Data Science II |
| 4 | Model Building and Assessment | - Intro to Statistical Models |
| 5 | Algorithms and Software Foundation | - Algorithms and Software Concepts<br>- Databases and Data Management |
| 6 | Data Curation | - Intro to Data Science I<br>- Intro to Data Science II<br>- Capstone Course |
| 7 | Knowledge Transference<br>-Communication and Responsibility | - Course in an Outside Discipline |

Berkeley’s implementation strategy for the DS program was based on starting early to build a data science mindset, no Berkeley’s pedagogical reflection signals that there is a challenge in maintaining student inclusion and diversity [4]. Students’ movement from one program to another might affect other program enrollment as well. Besides, the students are interested in enrolling upper division DS courses to continue with the DS program.

The DS program consists of a foundation of DS course, integrating concepts of computing, statics and critical thinking and connector courses that will lead to different domains. The starting point is a lower-division course on foundations of DS that was linked with a growing family of connector courses that tailor the program to a range of domains. The introductory courses make data science accessible to any Berkeley student relying only on university-entry mathematics requirements and no prior computing experience [4].

## IV. SYNERGY BETWEEN COMPUTER SCIENCE AND DATA SCIENCE

Although the journey of Data Science can be traced back to early 1960s [3], the latest surge of the interest in Data Science might have been the results of the NSF Big Data Initiative launched in 2012 and the new IEEE Transactions on Big Data inaugurated in 2015. Due to the fact that there are a significant portion of the overlap between Data Science and Computer Science based on the synergy built in the two, it is more effective and efficient to embed Data Science into Computer Science education, instead of bridging the two. In addition, as the investment in having a standalone program in Data Science is quite high, it is desirable to have Data Science concepts and skill introduced to CS majors in many budget-constrained institutions.

In the view of Data Science from all other areas, the bridging them with Data Science can be illustrated in Figure 3 below, which is reported by three major universities in New York University, UC Berkeley, and the University of Washington. As the Data Science Methodologies listed on the right hand side of the figure, all methods except Statistics is covered by Computer Science.

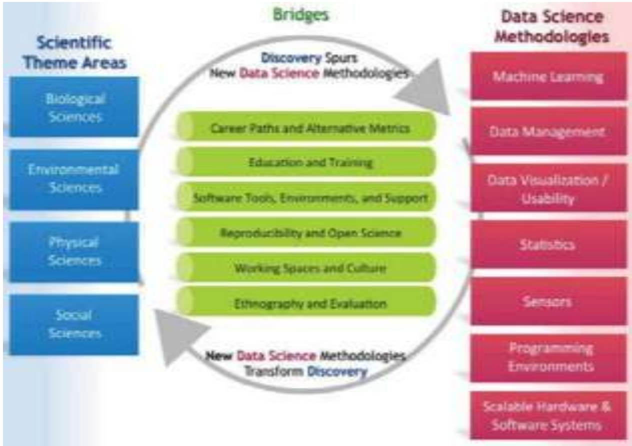


**Figure 3:** Bridging Data Science to Other Sciences [22]

The University of California at Berkeley developed and implemented undergraduate DS curriculum that is based on a modular-and-connector structure [4]. Faculty members from the Park City Math Institute (PCMI) 2016 [5] proposed a DS undergraduate curriculum based on six key competencies, where a DS program can be implemented using typical existing courses in Math, Computer Science, and Statistics as shown below in Table 5.

**Table 5:** Matching a DS program using typical existing course [5]

| Mathematics | Computer Science | Statistics |
|---|---|---|
| Calculus I | Introduction to CS | Introduction to Statistics |
| Calculus II | Data Structure/Algorithms | Statistical Modeling |
| Calculus III | Computer Architecture | Regression |
| Linear Algebra | Advanced Algorithms | Machine Learning |
| Probability Theory | Databases | Data Mining |
| Discrete Math | Software Engineering | Theory of Statistics |

Although “courses from the traditional disciplines of mathematics, statistics and computer science provide the basic infrastructure for the major at present” [5], those course in the three disciplines are required to have a well-planned redesign to integrate the elements needed for the competences to prepare our students for a promising career in Data Science as the authors summarized in [5].

CS2013 is the latest Curriculum Guideline for Undergraduate Degree Programs in Computer Science developed by ACM and IEEE-Computer Society. ACM and IEE-Computer Society update CS curriculum guideline regularly as to keep it modern and relevant. The CS2013[17] guideline presents updated body of knowledge, a result of rethinking the essentials necessary for a CS curriculum. The Knowledge Areas are categorized into a group of 18 Knowledge Areas corresponding to topical areas of study in CS. Table 6 below shows Data Science, Mathematics, and Statistics courses, which can be covered by the current undergraduate Computer Science curriculum, and those that cannot be covered based on CS2013 undergraduate curriculum guideline. This comparison is based on the authors’ understanding and analysis and used only for this study.

**Table 6:** Curriculum Comparison: Data Science vs. CS2013 Standard

| CS2013 Knowledge areas | Cover (47%) | Not Cover (53%) |
|---|---|---|
| • Algorithms and Complexity<br>• Architecture and Organization<br>• Computational Science<br>• Discrete Structures<br>• Graphics and Visualization<br>• Human-Computer Interaction<br>• Information Assurance and Security<br>• Information Management<br>• Intelligent Systems<br>• Networking and Communications<br>• Operating Systems<br>• Platform-based Development<br>• Parallel and Distributed Computing<br>• Programming Languages<br>• Software Development Fundamentals<br>• Software Engineering<br>• Systems Fundamentals<br>• Social Issues and Professional Practice | • Dataset Organization and Management<br>• Advanced Database Development using SQL<br>• Security Data Science<br>• Data Mining with Applications<br>• Applied Machine Learning<br>• Ethics<br>• Business Intelligence<br>• Information Systems and Data Management<br>• Big Data | • Introduction to Data Science<br>• Data Science Capstone<br>• Data Science Policy and Strategic Management<br>• Data Science for Everyone<br>• Intermediate Data Science<br>• Data Analytics<br>• Data Visualization<br>• Data Science Tools and Techniques<br>• Data Science Practicum<br>• Predictive Analytics |

The synergy between CS and DS can also be investigated based on Table 2 and Table 3 presented previously. As Table 7 shown below, the corresponding courses in Computer Science can be recommended along with the competences proposed in [5].

**Table 7:** DS vs. CS Courses for Each Competence

| | Competencies | Data Science Courses | Computer Science Courses |
|---|---|---|---|
| 1 | Computational Thinking | Algorithms and Software Concepts<br>Databases and Data Management | Algorithms; Software Engineering; Databases Management Systems; Data Warehouse; |
| 2 | Statistical Thinking | Intro to Statistical Models<br>Statistical and Machine Learning | Data Mining; Machine Learning; Artivicial Intellengence; |
| 3 | Mathematical Foundations | Mathematics for Data Science I<br>Mathematics for Data Science II | CS Math Requirements; |
| 4 | Model Building and Assessment | Intro to Statistical Models | Data Mining; Machine Learning; Artificial Intellegence; |
| 5 | Algorithms and Software Foundation | Algorithms and Software Concepts<br>Databases and Data Management | Algorithms; Software Engineering; Databases Management Systems; |
| 6 | Data Curation | Intro to Data Science I<br>Intro to Data Science II<br>Capstone Course | CS 0; CS 1; CS 2, Data Structure; Programming Languages; Capstone; HCI; Database Management System; Data Warehouse; |
| 7 | Knowledge Transference -Communication and Responsibility | Course in an Outside Discipline | Business<br>Healthcare<br>Biology |

Obviously, there is a synergy between Data Science and Computer Science as the facts that data itself is in digital, processed in digital, accessed in digital, and presented in digital. Computer Science is naturally the foundation of Data Science. The bottom line is that Data Science is so critical and so close to Computer Science that computer science students should be able to grasp it without too much effort. Data is the new "natural" resource for the next leap of the growth we cannot miss. It is not like Information Technology, Information Systems, or any other computer science related fields. This is the field Computer Science should take a significant part. Cars need gas or electricity to move while the society needs data to advance. Computer science students don't just learn how to use the collected data but study how to create and generate data and how to effectively and efficiently collect, store, and process data.

The core CF-DS includes common competences required for successful work of data scientist in different work environments in industry and in research and through the whole career path [6]. O'Reilly Strata industry research [8] defines the four data scientist profession profiles and their mapping to the basic set of technology domains and competencies as shown in Figure 4. The four profiles are defined based on the datascientists' practitioners' self-identification: 1) Data Business person; 2) Data Creator; 3) Data Developer; 4) Data Researcher.

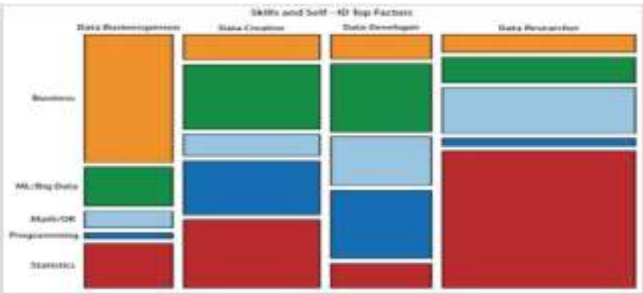

**Figure 4**: O'Reilly Strata survey on Data Scientist skills and profiles [8]

The related skills to each of the skill categories are defined in Figure 5 below:

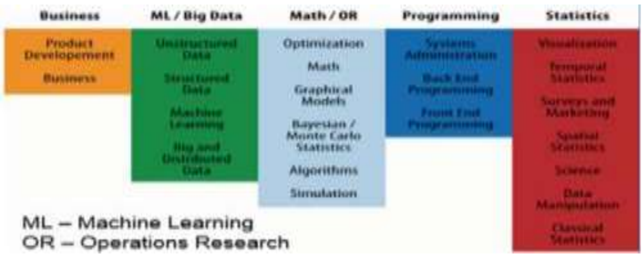


**Figure 5**: The skills under each of the skill categories [6]

Based on Figures 5 and 6 shown above, a computer science major better fits the skill sets as Data Developer and Data Creative. For a data researcher, more statistics knowledge is needed while for a data businessperson, the domain knowledge in a particular business plays a critical role.

As discussed above, there is a significant overlap between data science and computer science. It is possible for including data science concepts and techniques in computer science education without major reconstruction of the computer science curriculum.

## V. RECOMMENDATIONS

As discussed above, the goal of this paper is to see how to introduce DS theory and skill to CS students with a minimum disruption to CS curriculum. There is no one-size-fits-all solution for embedding DS into computer science education. The factors to consider when embedding DS into CS education are 1) levels of effort; 2) levels of resources; 3) levels of collaborations. In case of the levels of efforts, there has to be a multidisciplinary faculty committee to contribute in the program development with the management support. The main tasks of the committee can be the followings: 1) identifying and selecting the relevant computing and math courses from the existing computer science courses 2) developing new DS courses for the DS major which are not available within the CS department; 3) preparing the program course plan; 4) preparing course sequences; 5) preparing the program proposal.

By its multidisciplinary nature, a DS program should represent a combination of subject areas from several disciplines including applied mathematics, statistics, and computer science [14]. Consequently, schools that host DS programs are diverse: they include Business, Computer Science, Math, Mathematics and Statistics, and Arts and Sciences Schools.

Based on available resource, faculty preparation, curriculum development, and schedule management, recommended strategies for embedding DS into CS education are provided below in Table 8.

**Table 8:** Strategies for Embedding DS in CS Education

| | Carrier | Example |
|---|---|---|
| 1 | -Project<br>-Case Study | -Capstone Project<br>-Case Study in Software Engineering |
| 2 | -Course Module | -"Data Curation" in a Database Course<br>-"Scripting with R" in an OS course |
| 3 | -Regular Course | -Introduction to Data Science<br>-Data Analytics With R |
| 4 | -Multiple Courses (a Track or Concentration within a Major) | -Data Science I<br>-Data Science II |
| 5 | -Minor in Data Science | -A Core of Data Science Courses plus a Group of Selected Elective Courses |
| 6 | -Certificate in Data Science | -Similar to a Minor in DS but More Flexible in Admission and Completion |

It is not our purpose to provide a solution but a way of thinking in which everyone should be able to find their own way in introducing Data Science to their Computer Science students. Moreover, the strategies provided in Table 7 can be treated as a framework in which everyone can find a method in introducing DS into CS education no matter the levels of effort, resources, and collaborations they have. Due to the limited length of the paper, the discussion on how to implement those strategies is not included but will remain as part of our future work.

As studied in [9], more than 40% BSDS programs in the U.S. are offered by a Computer Science department or a department that offers a computer science program. Out of 10 areas for the future development

of data science as pointed out in [3], CS educators can well pursuit to "train the next-generation data scientists and data professionals who are qualified for data science problem-solving, with data literacy,
thinking, competency, consciousness, curiosity, communication, and cognitive intelligence, to work on the preceding data science agenda" [3].
Not just for Computer Science, some institutions want to embed Data Science to all majors, such as reported in [21] that Purdue University has kicked off a new initiative to "make data science education a part of every student's learning experiences on campus, no matter what field he or she is studying."

It was noticed that some institutions have moved their work further to bring Data Science to Computer Science education by blending the two into a single program in Computer and Data Science. The exemplary work can be found in [19] and [20].

## VI. CONCLUSION

Although Data Science is a promising field to explore new curricula, it is very challenging to have a standalone program for many computer science departments, especially those where the number of the students is under 100 and the number of the faculty members is under 5. As Data Science has many aspects as discussed above, it could be relatively easier for computer science educators to bring "data generation" and "data processing" through database related courses and "data analysis" through data mining and artificial intelligence courses. While embedding Data Science into Compute Science education requires tremendous effort from the faculty, our students, with the strength in Computer Science and the potential in Data Science, will be placed in a better position in facing the emerging challenges brought by this data embraced world.

## REFERENCES


[1] Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., Raghavan, P., Stodden, V., and Szalay, A. S., "Realizing the Potential of Data Science," the Communications of the ACM, April 2018, Vol. 61 No. 4, pp. 67-72

[2] Burtch, L., "The Burtch Works Study: Salaries of Data Scientists," May 2017, last accessed on Aug. 17, 2018 at https://www.burtchworks.com/wp-content/uploads/2017/05/DS-2017-Salary-Growth.pdf

[3] Cao, L., "Data Science: A Comprehensive Overview," the ACM Computing Surveys, Vol. 50, No. 3, Article 43, Publication Date: June 2017, pp. 43:1 to 43:42

[4] Carson, C., Culler, D., and Jacobsen, B., "Berkeley's Undergraduate Data Science Curriculum: Year 1 Pedagogical Overview," University of California, Berkeley, Nov. 2016, last accessed on Aug. 17, 2018, at http://data.berkeley.edu/sites/default/files/dsepsummaryreportyear1.pdf

[5] De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., Bryant,L., Cheng, L. Z., Francis, A., Gould, R., Kim, A. Y., Kretchmar, M., Lu, Q., Moskol, A., Nolan,D., Pelayo, R., Raleigh, S., Sethi, R. J., Sondjaja, M., Tiruviluamala, N., Uhlig, P. X., Washington, T. M., Wesley, C. L., White, D., and Ye, P., "Curriculum Guidelines for Undergraduate Programs in Data Science," the Annual Review of Statistics and Its Application, 2017. 4:2.1–2.16

[6] Demchenko, Y., Belloum, A., and Wiktorski, T., "EDISON Data Science Framework: Part 1. Data Science Competence Framework (CF-DS)," Release 2 v0.8, July 3, 2017, last accessed on Aug. 17, 2018, at http://edison-project.eu/sites/edison-project.eu/files/filefield_paths/edison_cf-ds-release2-v08_0.pdf

[7] Dumbill, E., Liddy, E. D., Stanton, J., Mueller, K., and Farnham, S., "Educating the Next Generation of Data Scientists," Big Data, Volume 1, Issue 1 / March 2013, pp. 21 - 27

[8] Harris, H. D., Murphy, S. P., and Vaisman, M., "Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work," O'Reilly Media, Inc., ISBN: 978-1-449-36824-1, June 2013: First Edition, March 2015: First Release: June 2013, Second Release, March 2015.

[9] Hassan, I. B. and Liu, J., "Data Science Academic Programs in the U.S.," Journal of Computing Sciences in Colleges, April 2019

[10] Louridas, P. and Ebert, C., "Embedded Analytics and Statistics for Big Data," IEEE Software, Volume 30, Issue 6, Nov.-Dec. 2013, pp. 33 - 39

[11] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H., "Big Data: The Next frontier for innovation, competition, and productivity," McKinsey Global Institute, May, 2011, last accessed Aug. 17, 2018, at https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

[12] Naimi, A. I. and Westreich, D. J., "Big Data: A Revolution That Will Transform How We Live, Work, and Think," Book Review, American Journal of Epidemiology, Volume 179, Issue 9, May 1, 2014, pp. 1143 - 1144

[13] National Institute of Standards and Technology, "NIST Big Data Interoperability Framework: Volume 1, Definitions," NIST Big Data Public Working Group, Definitions and Taxonomies Subgroup, Version 2,

June 2018, last accessed Aug. 17, 2018, at https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r1.pdf

[14] O'Neil, M., "As Data Proliferate, So Do Data-Related Graduate Programs," the Chronicle of Higher Education, February 3, 2014, last accessed Aug. 17, 2018, at https://www.chronicle.com/article/As-Data-Proliferate-So-Do/144363

[15] Parks, R. F., Ceccucci, W., and McCarthy, R., "Harnessing Business Analytics: Analyzing Data Analytics Programs in U.S. Business Schools," Information Systems Education Journal, 16(3), June 2018, pp. 15 - 25

[16] Tang, R. and Sae-Lim, W., "Data science programs in U.S. higher education: An exploratory content analysis of program description, curriculum structure, and course focus," Journal of Education for Information, Volume 32, No. 3, July 28 2016, pp. 269 - 290

[17] The Joint Task Force on Computing Curricula, "Curriculum Guidelines for Undergraduate Degree Programs in Computer Science," Association for Computing Machinery (ACM) and IEEE Computer Society, December 20, 2013

[18] The National Academies Press, "Data Science for Undergraduates: Opportunities and Options," A Consensus Study Report of The National Academies of Sciences, Engineering, and Medicine, 2018

[19] University of Saint Joseph, "BS in Computer and Data Science Program," last accessed Aug. 22, 2018, at https://www.usj.edu/academics/schools/sasbe/mathematical-sciences/computer-and-data-science-bs/

[20] Lebanon Valley College, "BS in Computer and Data Science Program," last accessed Aug. 22, 2018, at http://www.lvc.edu/academics/programs-of-study/computer-data-science/degree-requirements/

[21] Schaffhauser, D., "Purdue to Embed Data Science into Every Major," Campus Technology, April 19, 2018, last accessed Aug. 25, 2018, at https://campustechnology.com/articles/2018/04/19/purdue-to-embed-data-science-into-every-major.aspx

[22] The Moore-Sloan Data Science Environments, "Creating Institutional Change in Data Science," New York University, UC Berkeley, and the University of Washington, last accessed Aug. 25, 2018, at http://msdse.org/files/Creating_Institutional_Change.pdf

[23] Data Science Center, "Ph.D. program in Data Science," New York University, last accessed Aug. 27, 2018, at https://cds.nyu.edu/academics/phd-in-data-science/

[24] Analytics and Data Science Institute, "Ph.D. program in Analytics and Data Science," Kennesaw State University, Atlanta, Georgia, last accessed Aug. 27, 2018, at http://graduate.kennesaw.edu/datascience/

[25] Department of Statistics, "Bachelor Program in Data Science," University of California Irvine, last accessed Aug. 27, 2018, at https://www.stat.uci.edu/slider/b-s-in-data-science/

[26] Department of Mathematics and Statistics, "Bachelor Program in Data Science," the University of Vermont, last accessed Aug. 27, 2018, at https://www.uvm.edu/cems/mathstat/bs_data_science

[27] Department of Computer Science, "Data Science, B.S.," Northern Kentucky University, KY, last accessed Aug. 30, 2018, at http://nku.catalog.acalog.com/preview_program.php?catoid=9&poid= 1340

[28] Department of Computer Science, "Data Science, Bachelor of Science," Salem University, WV, last accessed Aug. 30, 2018, at https://www.salemu.edu/degrees/bachelors-data-science/