

Undergraduate Degree Programs in Data Science

Debatosh Debnath
Dept. of Computer Science and Engineering
Oakland University, Rochester
Michigan 48309, U.S.A.

Abstract

To meet the growing demand for well-trained data science graduates, institutions of higher education have started to offer data science programs at the undergraduate level in recent years. In this paper, we show that such programs can be implemented in many campuses by developing at most five new courses.

1 Introduction

A report from the McKinsey Global Institute in 2011 stated that the volume of worldwide data is increasing at a rate of approximately 40 percent per year and the United States alone will have a shortage of up to 190,000 data scientists and 1.5 million managers with data analytical skills by 2018 [1].

To fill the growing demand for data science professionals, our higher education institutions are responding by creating new initiatives and degree programs in data science [2, 3]. In Michigan alone two new undergraduate programs in data science have been introduced in fall of 2015—one at the University of Michigan and the other at the Western Michigan University. College of Charleston, University of Rochester, University of Massachusetts Dartmouth, University of Nebraska Omaha, University of San Francisco, Ohio State University, Northern Kentucky University, and Ottawa University are some of the other institutions that have developed undergraduate degree programs in data science or data analytics in recent years. Data science has become so important at our national level that the U.S. government appointed its first ever Chief Data Scientist in 2015.

The key focuses of any data science program at the undergraduate level lie in mathematics, statistics, and computing. Mathematics helps students develop reasoning and problem-solving skills, and statistics and computing provide the necessary analytical skills and tools to represent data in a meaningful way. This paper summarizes the curricula of several undergraduate programs in data science and provides an outline of core contents that every data science program should cover. We also show that a data science program at the undergraduate level can be implemented by developing several new courses and selecting appropriate courses from undergraduate programs in mathematics, statistics, and computer science.

2 Overview of Several Data Science Programs

University of Michigan

The University of Michigan introduced a new undergraduate degree program in Data Science in fall 2015. The program is jointly managed by the Division of Computer Science and Engineering in the College of Engineering (CoE) and the Department of Statistics in the College of Literature, Science, and the Arts (LSA). Students from the both colleges can declare a major in data science after completion of a set of prerequisite courses. CoE students receive a BSE degree in Data Science and LSA students receive a BS degree in Data Science. The requirements for BSE and BS degrees are slightly different. In the following we focus on the requirements for the BSE degree only.

Students in CoE are required to complete 55 credits of mathematics, physics, chemistry, engineering, computer science, and intellectual breadth courses that come from humanities, liberal arts, and professional and creative development areas. The program core is 30 credits and includes discrete mathematics, computer programming, data structures, computer algorithms, probability and statistics, applied statistics, database management and applications, either machine learning or data mining, and a data science applications elective which can be an upper-level course related to applications of data science. There are over 40 pre-approved data science applications elective courses.

There are 11 credits of flexible technical electives, 8 credits of advanced technical electives, 15 credits of general electives, 4 credits of capstone experience, and 5 credits of technical communication and professionalism courses. Flexible technical electives are pre-approved 200- or higher-level courses and come from at least 24 different areas of science, engineering, mathematics, statistics, business, and arts. Flexible technical electives can be used to satisfy pre-requisites for advanced technical electives or to provide foundation for the data science application elective. Advanced technical electives for data science are a set of over 30 pre-approved 300- or higher-level courses; they are primarily courses in statistics, computer science, and their very closely related areas. Students can choose a capstone course from a pool of at least nine courses. General electives allow students to explore any area of intellectual endeavor with some restrictions.

Western Michigan University

Western Michigan University also started a BS degree program in Data Science in Fall 2015. The program is a joint effort of the Department of Computer Science and Department of Statistics.

Students in this program take 8 credits of mathematics courses as background support. There are 15 credits in computer science core which includes courses in programming for data science, storage and retrieval of big data, machine learning, and senior design. Students also take a 16-credit statistics core which includes courses in statistics using a programming language, statistical computing, regression analysis, applied data mining, and computer-based data analysis. There are

15 elective courses from statistics and computer science, and the students are required to choose two courses that add 6 more credits towards the data science degree. Students also take one or more courses in software systems development as a capstone project that satisfies the writing requirement. Students take additional courses to satisfy general education and other departmental as well as university requirements.

University of Rochester

The University of Rochester offers both BS and BA degrees in Data Science. After completing four pre-requisite courses in mathematics and computer science, students at the University of Rochester take eight courses to satisfy the data science core requirement. The core includes courses in mathematics, statistics, data mining, artificial intelligence, database systems, design and analysis of algorithms, and data science capstone. There are multiple statistics courses in the core and students are required to take only two of them. Students are also required to take three upper-level courses from an application area. There are opportunities to select biology, brain and cognitive science, computer science, mathematics and statistics, earth and environmental science, physics, economics and business, or political science as the application area.

Students interested in receiving a BS degree in Data Science take an additional three courses, two of which are from probability and statistics and one from computer science. The computer science courses include machine learning, machine vision, natural language processing, artificial intelligence, and computer organization.

Northern Kentucky University

Northern Kentucky University introduced a BS degree program in Data Science in Fall 2013. As part of the program, students take 18 credits of mathematics and statistics courses that include three calculus courses, a course in linear algebra, and two courses in probability and statistics. Students are also required to take 15 credits of computer science courses which include two courses in object-oriented programming, a course in data structures and algorithms, and a course in database management. There are 16 credits of required data science courses such as data analytics, data mining, data visualization, big data, and data science capstone. Students also complete a 3-credit course in microeconomics as well as 15 credits from courses in the area of business informatics such as business process analysis, management information systems, IT project management, business analytics, and workflow design and management. Students also take 9 credits of upper level courses from either a computation-intensive track or business analytics track. There are a set of courses to choose from for each track.

University of San Francisco

The University of San Francisco offers an undergraduate degree in Data Science. Students in data science complete 56 credits of base curriculum to gain an array of mathematical, statistical, and computational skills. Mathematical and statistical courses include four courses in calculus, linear

algebra, and discrete mathematics as well as three courses in probability, statistics, and mathematical modeling. Computer programming, data structures, data visualization, and data mining provide computational skills. Students take a senior project or an applied mathematics laboratory course to satisfy the capstone experience. A course in economics is also part of the base curriculum. In addition, students complete 12 credits from one of the three concentration areas, *viz.*, computational data science, mathematical data science, and economic data science. Students under computational data science take courses in software development, databases, and numerical analysis. Mathematical and economic data sciences students are required to take upper-level courses in mathematics and economics, respectively. Students also take additional courses to satisfy university requirements.

3 Core Courses in Data Science Programs

A close look at the programs outlined in the previous section reveals that there is a general consensus on the core contents of any undergraduate level data science program. To develop the necessary reasoning, problem-solving, and analytical skills, students in data science should take courses in calculus, linear algebra, probability and statistics, discrete mathematics, computer programming, and data structures. Students majoring in data science should take a course in a data science oriented programming language such as R or Python. Courses in database management, machine learning and data mining, data visualization, computer-assisted data analysis, and applied statistics should be part of any data science curriculum to add the necessary analytical skills and tools. Students in data science should also take courses from an application area, including but not limited to domains such as biology, business, engineering, and health science. In addition, students must be prepared for the practice of data science through a capstone course that provides the opportunity to apply the skills and tools learned in earlier course work to an application area.

The majority of courses for the data science programs mentioned above can be selected from existing degree programs in mathematics, statistics, and computer science. Courses in calculus, linear algebra, probability and statistics, and applied statistics are available as part of almost all mathematics and statistics undergraduate programs. Similarly, almost all computer science undergraduate programs have courses in discrete mathematics, computer programming, data structures, and database management systems. Some campuses already offer courses in machine learning and data mining, data visualization, and computer-assisted data analysis as part of either statistics or computer science programs. Almost all the campuses have plentiful courses in application areas.

4 Conclusions

Our analysis shows that the majority of required courses for a data science program at the undergraduate level can be selected from existing programs in mathematics, statistics, and computer science. To implement a data science program, in addition to developing a capstone course in

data science, many campuses are at most required to develop courses in a data science oriented programming language such as R or Python, machine learning and data mining, data visualization, and computer-assisted data analysis.

References

- [1] Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, June 2011.
- [2] P. Anderson, J. McGuffee, and D. Uminsky, “Data science as an undergraduate degree,” in *Proc. 45th ACM Technical Symposium on Computer Science Education*, pp. 705–706, Mar. 2014.
- [3] U-M Launching \$100 Million Data Science Initiative, Sept. 2015. [Online] Available: <https://record.umich.edu/articles/u-m-launching-100-million-data-science-initiative>. [Accessed: 9-Feb-2016]