A Transdisciplinarity Approach to an Undergraduate Degree in Data Science

James McGuffee, Qi Li, and Joe Rubleske
College of Informatics, Northern Kentucky University

ABSTRACT

This position paper describes the creation and implementation of the bachelor’s degree in Data Science at Northern Kentucky University (NKU). In addition to being one of the few undergraduate degree programs in Data Science in the United States, the Data Science program at NKU is unique in that it was created with transdisciplinarity in mind requiring full cooperation from three existing academic departments: Business Informatics, Computer Science, and Mathematics and Statistics.

TRANSDISCIPLINARITY and BIG DATA

In 1987, German scientific philosophy writer Jürgen Mittelströsser introduced the concept of transdisciplinarity. Going beyond interdisciplinary cooperation, the term “…transdisciplinarity is intended to imply that cooperation will lead to an enduring and systematic scientific order that will change the outlook of subject matters and disciplines [4].” In other words, transdisciplinarity involves the investigation of a subject or problem through the lens of two or more disciplines because the problem or subject being examined is beyond the scope of any one academic discipline. It’s important to note that transdisciplinarity is not the creation of a new discipline. For example, computer science is a relatively new academic discipline that has contributions from other disciplines but over the past half century has established its own discipline specific artifacts and methodologies. In contrast, a transdisciplinarity approach doesn’t try to create a new discipline but seeks to utilize the advantages of all the disciplines that come together to examine the subject or solve a problem.

Why approach data science (a.k.a. the study of big data) from a transdisciplinarity perspective? When working with big data there are many proposed steps that are needed to glean maximal value. These steps include data acquisition, data cleaning, information extraction, data integration, modeling, analysis, interpretation, and deployable actions based on big data analysis [3]. Due to the constraints of a single discipline, many discussion of big data falter by focusing on only one or two steps. Additionally, with each step come a set of challenges. Some of the big challenges in big data include the heterogeneity of the data being collected, the timeliness of the data, privacy concerns, appropriate data visualizations, and the large number of diverse hardware and software tools that exist to manage and explore big data [3]. The problems and challenges associated with the study of and research in big data seem ripe for a transdisciplinarity approach.

Northern Kentucky University has developed a data science degree program from a practical and pragmatic transdisciplinarity centered approach [1]. This program formally began in the fall of 2013 and has garnered much interest from both internal and external communities. The next four sections of this paper will describe the details of the data science degree plan at NKU. Before we describe the unique data science courses, let’s first examine the individual contributions of each of the three cooperating departments.
MATHEMATICS AND STATISTICS

As part of the data science degree, students are required to take nine hours of calculus, six hours of probability and statistics, and three hours of linear algebra. With a minimum of eighteen credit hours in mathematics and statistics, the data science graduate should have a solid background in the quantitative knowledge and skills needed to successfully work with big data. In addition to the required courses in the degree plan, a data science student must take nine hours of guided electives. As part of the guided elective requirements, there are six courses from mathematics and statistics that a student may choose from. The three guided elective courses from mathematics are differential equations, advanced calculus, and applied mathematical models. The three guided elective courses from statistics are regression analysis, time series analysis, and an advanced probability course.

BUSINESS INFORMATICS

Students enrolled in the data science degree program must take eighteen hours of business informatics courses. These courses include a principles course, business process analysis, management information systems, IT project management, quantitative analysis, and workflow design and management. In addition to the required business informatics courses, the data science student must also take the principles of microeconomics course from the Economics Department. The data science guided electives in business informatics include information systems analysis, information security, and business intelligence applications.

COMPUTER SCIENCE

The core courses from computer science include nine hours of computer programming in Java, six hours in database management systems, and three hours in computer security. The guided electives from computer science include web programming, computer systems, advanced programming methods, artificial intelligence, and the design and analysis of algorithms.

DATA SCIENCE

There are three core faculty members in the data science program at Northern Kentucky University. There is a faculty member from each of the three cooperating academic departments. In addition to teaching and research in their respective home disciplines, these faculty members are also responsible for the creation, management, and teaching of six required data science specific courses. The six courses are:

- DSC 101 Introduction to Data Science
- DSC 311 Data Analytics
- DSC 321 Data Visualization
- DSC 411 Data Mining
- DSC 421 Big Data
- DSC 496 Data Science Capstone
DSC 101 is a seminar course for first semester freshmen that describes the field of data science, explores career possibilities, and gives hands on experiences with commonly used tools. The junior level data analytics course, DSC 311, covers the data analysis process, data modeling techniques, the identification of important attributes, and the presentation of data analysis results. The data visualization course, DSC 321, includes using the Griffin Hall Digitorium Christie digital micro tile wall as a laboratory for visualization [2].

The course in data mining, DSC 411, is concerned with discovery of patterns in large datasets. Topics in this course include classification, clustering, associations, anomaly detection, selecting appropriate data mining algorithms, and the use of data mining tools, such as SPSS, R, Weka, or Oracle Data Miner. The big data course, DSC 421, covers the manipulation, storage, and analysis of large scale data, the use of a large-scale distributed file system (e.g. Hadoop Distributed File System), the use of large-scale databases, and MapReduce algorithm design. The capstone course, DSC 496, is a student projects course that includes the cleaning, processing, and analysis of data, along with oral and written presentations with appropriate data visualizations.

In addition to the six required courses, there is currently one additional data science course that may be used as a guided elective. DSC 431 is a networks analysis course that covers various types of networks: computer, social, and information.

IMPLEMENTATION

There are three core faculty members with responsibilities for the data science degree program. These faculty members are Dr. Mark Lancaster from Mathematics and Statistics, Dr. Qi Li from Computer Science, and Dr. Joe Rubleske from Business Informatics. At the current time there is also an open position for a lecturer in Data Science who will also coordinate and manage student data analytics projects in the College of Informatics’ Center for Applied Informatics (CAI).

The degree plan officially began in the fall of 2013 with formal approval from Kentucky’s Council on Postsecondary Education (CPE). For the first section of DSC 101 (fall 2013) there were 10 students. This year (fall 2014) there are 21 new students enrolled in the DSC 101 course. The DSC 311 course will be offered for the first time in spring 2015. We expect our first graduates in this program in May of 2017.

Our expectation is that the graduates of the Data Science program at Northern Kentucky University will be able to understand the mathematical and statistical foundations of data science and understand the business context in which data science functions. We also expect students to be able to implement algorithms for data aggregation, cleaning, and analysis, select and apply appropriate data analysis techniques to a variety of tasks, and communicate data analysis findings with appropriate visualizations.
SOURCES

1. Bachelor of Science in Data Science, website <http://datascience.nku.edu>.

ABOUT THE AUTHORS

- Dr. James McGuffee is the Chair of the Department of Computer Science at NKU. The data science degree program is administrated through the Computer Science Department.
- Dr. Qi Li is an Assistant Professor of Data Science. She is also a faculty member in the Department of Computer Science.
- Dr. Joe Rubleske is an Assistant Professor of Data Science. He is also a faculty member in the Department of Business Informatics.