

The University Senate of Michigan Technological University

Proposal 20-14

(Voting Units: Academic)

“Proposal for a New Non-Departmental Graduate Certificate in Data Science”

February 28, 2014

Contacts: Laura Brown (Computer Science), Mari W. Buche (School of Business & Economics), Gowtham S (Information Technology Services), Timothy Havens (Electrical & Computer Engineering/Computer Science), Jacqueline Huntoon (Graduate School), Saeid Nooshabadi (Electrical & Computer Engineering/Computer Science), and Allan Struthers (Mathematics)

e-mail: datascience@mtu.edu

Executive Summary

This proposal describes a plan for a new Graduate Certificate in Data Science. This program will augment the proposal for a Master of Science (M.S.) in Data Science. Like its M.S. counterpart the Graduate Certificate in Data in Science the has three main objectives: i) to attract students from various disciplines who wish to learn the basics of data analysis, data science, and computing tools; ii) to teach students basic skills in communication and build their awareness of business contexts; and iii) to provide students the opportunity to gain the basic skills that give them the ability to analyze large data sets, including Big Data.

Our goal is to have all the core data science courses and most approved data science courses offered online by 2016, which we note would allow off-campus students to fully complete the Graduate Certificate in Data Sciences with online offerings.

1. Background

The proposed Graduate Certificate is offered as subset of M.S. in Data Science and is offered to science and engineering graduates who wish to upgrade their qualification to be able to work in a profession with a primarily role to manage and analyze data.

2. Justification and Estimated Market

The program we are proposing will significantly increase the number of data scientists that Michigan Tech can offer to the workforce. The Graduate Certificate program in Data Science like its M.S. counterpart will provide students with strong academic training in data analysis in a range of areas (e.g., physical sciences, geosciences, geoinformatics, bioinformatics, cheminformatics, environmental, social sciences, business and commerce) while at the same time introduce essential business acumen, communication and teamwork skills highly valued by industry and government.

The proposed program emphasizes data analytics from a general perspective, but the skills to be learned are applicable to a diverse range of areas, including business analytics, computer science and engineering, and informatics. To support the interdisciplinary nature of the Data Science program, applications from multiple areas will be included in the coursework.

The proposal Data Science program is in line with [Michigan Tech strategic plan](#)¹ to “be a leader in creating solutions for society's challenges through education and interdisciplinary endeavors that advance sustainable economic prosperity...”

3. Competitive Analysis

Established computer science, business analytics, and statistics master's degrees and certificate programs already exist, both in the U.S. and abroad, and provide specializations in data mining and predictive analytics. However, despite interest and recognized need, there are as yet only a few programs dedicated to data science in the U.S. Further, the existing programs have been designed around business data with a less domain-specific scientific focus. These master's programs include Northwestern's new M.S. in Analytics, DePaul's M.S. in Predictive Analytics, University of San Francisco's M.S. in Analytics, LSU's M.S. in Analytics, Rutgers's Professional Science Master's (PSM) of Business and Science in Analytics, and NCSU's M.S. in Analytics (also a PSM program).

Finally, there is increased recognition by federal agencies that supporting Big Data research is important. For example, the National Institutes of Health (NIH) director, Dr. Francis Collins, recently convened a “Data and Informatics Working Group” that made several key recommendations aimed at fostering NIH sponsored research in Big Data. Other federal agencies have also signaled interest in Big Data research, including National Science Foundation, DARPA, Department of Energy, and Department of Defense.

¹ STRATEGIC PLAN https://www.banweb.mtu.edu/pls/owa/strategic_plan2.p_display

4. Detailed Description of Graduate Certificate in Data Science

i. Title:

Graduate Certificate in Data Science

ii. Catalog description:

The non-departmental Data Science program at Michigan Tech provides a foundation for the emerging field of “Big Data” science, including the use of data mining, predictive analytics, cloud computing, and business skills, with a domain specific specialization. The main threads of analytic

techniques, programming practice, domain knowledge, business acumen, and communication skills are intertwined in this program.

The Graduate Certificate in Data Sciences provides the basic skills in data analytics, data management, business and communication skills. Entry into this program assumes basic knowledge in statistical and mathematical techniques, programming, and communications.

iii. Credits:

Graduate Certificate in Data Science 15 credits (minimum)

iv. Course work:

In accordance with Senate policy, the requirements for the interdisciplinary Graduate Certificate in Data Sciences are a minimum 15 credits of coursework, including the required 12 credits of core courses and 3 credits of approved Data Science electives. All other requirements are per Senate proposals 11-10 and 4-11. Because this is an interdisciplinary certificate, a maximum of six credits can be earned at the 3000-4000 level.

Coursework Summary

Core courses for M.S. Data Science (12 credits):

The four required core 3-credit courses focus on basic skills in data science analytics, data mining, and business analytics. These courses are:

- UN 5550 - Introduction to Data Science (3 credits)²
- MA 4790 - Predictive Modeling (3 credits)
- CS 4821 / MA 4795 - Data Mining (3 credits)
- BA 5200 - Information Systems Management and Data Analytics (3 credits)³

² New course to be designed for Fall 2014; submitted to curriculum proposal (binder) Fall 2013. This course will be administered by the Office of Dean of the Graduate School, and executed by Data Science faculty across the campus.

³ Revise course BA 5200-Strategic IS Management; submitted to curriculum proposal (binder) Fall 2013

Approved Data Science elective courses for M.S. Data Science (minimum of 3 credits):

The remaining 3 credits for the graduate certificate must be taken from the approved 3-credit Data Science elective courses that are as part of the M.S. program in Data Science as below:

- CS 5841 / EE 5841 - Machine Learning (3 credits)⁴
- CS 5491 - Cloud Computing (3 credits)⁵
- CS 5471 – Advanced Topics in Computer Security (3 credits)⁶
- MA 5781 - Time Series Analysis and Forecasting (3 credits)⁷
- BA 5740 - Managing Innovation & Technology (3 credits)
- PSY 5210 - Advanced Statistical Analysis and Design I (4 credits)
- FW 5083 - Bioinformatics Programming and Skills (3 credits)⁸

Foundational Prerequisite Requirement:

It is expected that students seeking enrollment in this program will have sufficient foundational skills and aptitude in computer programming, statistical analysis, information systems and databases. The required foundational skills may have been obtained through formal academic qualifications, work experience, or a combination. Students will be encouraged to develop their foundational skills before coming to Michigan Tech to start the graduate certificate program in Data Science. After taking the entrance assessment exam and evaluation of the student's application, students will receive advice regarding their skill competence and may be required to take specific foundational courses (Appendix I) as necessary to acquire the required level of foundational skills. As students matriculate in the program, their assigned advisors will continually monitor students' progress to ensure that students are given all the necessary advice that they need to be successful in the program. Appendix I provides a list of foundational courses.

v. Online delivery:

Our goal is to have all the core data science courses and most approved data science courses offered online by 2016, which we note would allow off-campus students to fully complete the Graduate Certificate in Data Sciences with online offerings. Note that BA 5200 - Information Systems Management and Business Analytics will be offered as an online course starting in 2014. Additionally, the approved Data Science courses, CS 5841 / EE 5841 - Machine Learning and CS 5491 - Cloud Computing, will be offered as online courses in 2016.

⁴ New course to be designed for Spring 2015; submitted to curriculum proposal (binder) Fall 2013

⁵ New course to be designed for Spring 2015; submitted to curriculum proposal (binder) Fall 2013

⁶ New course to be designed for Spring 2015; submitted to curriculum proposal (binder) Fall 2013

⁷ Graduate version of MA 4780, submitted to curriculum proposal (binder) Fall 2013, to be offered as a split-level undergraduate/ graduate course. The graduate version of this course contains additional theoretical material and substantial project work.

⁸ Graduate version of FW 4099, submitted to curriculum proposal (binder) Fall 2013, to be offered as a split-level undergraduate/ graduate course. The graduate version of this course contains additional theoretical material and substantial project work.

vi. Description of new or revised Data Science courses (offered as part of M.S. in Data Science program):

All new (or revised) courses were added (modified) in the curriculum proposal (binder) process of Fall 2013.

UN 5550 - Introduction to Data Science (new) (3 credits)

This course provides an introduction to Big Data concepts, with focus on data management, data modeling, visualization, security, cloud computing, and data science from different perspectives: computer science, business, social science, bioinformatic, engineering, etc. This course also introduces the tools for data analytics such as SPSS Modeler, R, SAS, Python, and MATLAB. It involves two case study projects, each of which is integrated with communication and business skills.

BA 5200 - Information Systems Management and Data Analytics (revision) (3 credits)

BA 5200 Focuses on management of Information Systems /Information Technology within the business environment. Topics include Information Technology infrastructure and architecture, organizational impact of innovation, change management, human-machine interaction, and contemporary management issues involving data analytics. Class format includes lecture, group discussion, and integrative case studies.⁹

CS 5841 / EE 5841 - Machine Learning (new) (3 credits)

This course will explore the foundational techniques of machine learning. Topics are pulled from the areas of unsupervised and supervised learning. Specific methods covered include naive Bayes, decision trees, support vector machines (SVMs), ensemble, and clustering methods.

CS 5471 - Advanced Topics in Computer Security (new) (3 credits)

This course covers various aspects of producing trusted computer information systems.

Topics may vary; network perimeter protection, host-level protection, authentication technologies, formal analysis techniques, and intrusion detection will be emphasized. Current systems will be examined and critiqued.

⁹ Expanded description of BA 5220: This course is a restructuring of the existing course BA 5200 - Strategic IS Management to achieve a more acute focus on data analytics. The course incorporates experiential application of methods and analysis of business case studies focusing on contemporary issues in data analytics (i.e., Big Data) to include comprehension of business and organizational context, visualization and interpretation of results, reporting of outcomes from data analytics, evaluation of alternative techniques, and other current topics. Multiple online resources will be employed, including Teradata University. Students in this class will utilize open source software (e.g. Hadoop and NoSQL), developing skills applicable to industry. Ethical foundations and managerial constraints will be integrated throughout the course

CS 5491 - Cloud Computing (new) (3 credits)

This course provides an overview of the principles, methods, and leading technologies of cloud computing technologies. Topics include cloud computing concepts and architecture: Hadoop, MapReduce; standards; implementation strategies; Software as a Service (SaaS); Platform as a Service (PaaS); Infrastructure as a Service (IaaS); workload patterns and resource management; migrating to the cloud; and case studies and best practices. Students in this class will build their own cloud application using services from providers such as Amazon or IBM.

5. Estimated Costs For Financial Evaluation

The Graduate Certificate in Data Science program is a subset of M.S. in Data Science program, and does not incur any cost in addition and beyond the M.S. program. The approval of Graduate Certificate program must be subject to the approval of the M.S. program.

6. Planned Implementation Date

This program has an anticipated start in Fall semester, 2014. This program will be offered as a regular program. The program will be extended into an online program as soon as it is established and practical to do so. We envision a start date of Fall 2016 for the online delivery of this program.

7. Program Governance

Like other non-departmental and interdisciplinary programs at Michigan Tech, the Data Science program will be administered through the Graduate School, which will have the overall responsibility and final oversight for the program. The program will have the same management structure that governs the M.S. in Data Science program.

Appendix I: Foundational Skills Courses

It is expected that students seeking enrollment in this program will have sufficient foundational skills and aptitude in computer programming, statistical analysis, information systems and databases.

For those students who need additional training in these areas, the courses listed below will help build skills necessary for successful completion of the certificate. Not all students will need to take these courses as such the foundational courses are not required. Note, for students coming from a Bachelor's program at Michigan Tech, the foundational courses do not "double-count" for both the B.S/B.A. program and the Graduate Certificate in Data Science.

Note that 2000 level courses listed here cannot be counted towards the requirement for the Graduate Certificate in Data Science degree, but may be necessary for a given student to build their foundational knowledge.

- MA 2330 - Introduction to Linear Algebra (Credits: 3)
- MA 3710 - Engineering Statistics (Credits: 3)
- MA 3715 - Biostatistics (Credits: 3)
- MA 3740 - Statistical Programming and Analysis (Credits: 3)
- MIS 2000 - IS/IT Management (Credits: 3)
- MIS 2100 - Introduction to Business Programming (Credits: 3)
- MIS 3100 - Business Database Management (Credits: 3)
- MKT 3600 - Marketing Research (Credits: 3)
- CS 2321 - Data Structures (Credits: 3)
- CS 3425 - Database (Credits: 3)
- SAT 3002 - Application Programming Introduction (Credits: 3) ⁹
- SAT 3210 - DB Management (Credits: 3) ¹⁰
- SAT 4600 - Web Application Development (Credits: 3) ¹¹

⁹ New 3-credit course designed for Fall 2014

¹⁰ Summer offerings available.

¹¹ New 3-credit course designed for Spring 2015

Introduced to Senate: 05 March 2014

Approved by Senate: 26 March 2014

Approved by Administration: 03 April 2014