The University Senate of Michigan Technological University

Proposal 19-14

(Voting Units: Academic)

"Proposal for a New Non-Departmental Master of Science 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

The proposed Master of Science (M.S.) in Data Science will be the first non-departmental master's degree at Michigan Technological University. The M.S. Data Science has three main objectives: i) to attract students from various disciplines who wish to learn the key concepts of data analysis, data science, and computing tools; ii) to teach students necessary skills in communication and build their awareness of business contexts; and iii) to provide students the opportunity to gain domain specific skills that give them the ability to analyze large data sets, including Big Data.

The M.S. program will be developed to adhere to national requirements for Professional Science Master's[1] (PSM) programs, where the emphasis is on advanced training in science and engineering, while simultaneously developing highly-valued business and communications skills. Once the M.S. program is approved at Michigan Tech it will be submitted to the national PSM oversight organization for accreditation as a PSM. A plan to offer the M.S. as an accelerated master's will also be developed following approval of the program by Michigan Tech.

---

[1] http://www.sciencemasters.com/

1. Background

The Internet has steadily moved from text-based communications to richer content, including interactive maps, images, videos, and most importantly metadata such as geolocation information and time and date stamps. High-speed communication networks such as 3G, 4G, and WiFi have enabled fast transmission of these storage-intensive data. The amount of data captured by e-health networks,

telematics and telemetry devices for monitoring the location, movements, status of mobile units, for use in machine-to-machine and people-to-machine systems, social networks, environmental agencies, commercial and business agencies, and security agencies is exploding. In the year 2000, the amount of data stored in the world was about 800,000 petabytes (one petabyte = one million gigabytes). This amount is expected to reach 35 zettabytes (one million-million petabytes) by 2020. Twitter and Facebook, respectively, generate more than 7 terabytes of data each day. Advances in data storage and data-mining technologies make it possible to preserve increasing amounts of data generated directly or indirectly by users.

As we stand at a point where our economy is driven by Big Data, our data collecting abilities have far outpaced techniques to manage and analyze these data. Hence, enhanced capabilities in data analysis are needed to obtain valuable new insights from these captured data. Examples are sensor networks, big social data and social networks analysis, telephone call meta-data, military surveillance, medical records, imaging and video archives, large-scale e-commerce, astronomy, atmospheric science, genomics, biogeochemical, biological, and other complex and often interdisciplinary scientific research.

The field of data science has emerged as a response to increased data abundance in industry, science, and engineering. The NationalConsortiumforDataScience[2] (NCDS), a collaboration of industry and academic institutions, was formed to identify data science challenges, coordinate research priorities, to support the development of technical and ethical data standards policy, and to foster economic growth by launching a national strategic
initiative to secure the U.S. as the world leader in data science.

The Big Data explosion needs data scientists and analysts able to interpret massive data sets. The lack of trained data scientists has meant that less than 5% of data are used effectively, according to the Forresterresearchfirm[3].

Data scientists primarily manage and analyze data, which requires computer science (CS), statistics, business, marketing, and communications skills. Traditional statistics training lacks the emphases on required CS and domain-specific skills, while traditional CS and engineering training lack emphases on the required statistical analyses skills. Furthermore, both lack acumen in business, marketing, and communications. Data analysis also requires expertise in the specific domain of the application (e.g., engineering, imaging and video analytics, social sciences, bioinformatics, etc).

[2] http://data2discovery.org
[3] BIG DATA WILL HELP SHAPE YOUR MARKET'S NEXT BIG WINNERS http://blogs.forrester.com/brian_hopkins/11-09-30-big_data_will_help_shape_your_markets_next_big_winners

2. Justification and Estimated Market

A simple job search for "Data Scientist" today reveals thousands of job openings. The Bureau of Labor Statistics (BLS) forecasts a 19% growth in employment for computer and
information research scientists by 2020. Numerous articles, studies, and blog postings warn of the shortage of Data Scientists, e.g., InformationWeek[4]; Fortune[5]; and EMCDataScience Study[6]. In 2011, McKinsey Global Institute published "Bigdata[7]: The next frontier for innovation, competition, and productivity," citing a need for 140,000-190,000 data scientists in the U.S. alone, by 2018.

The program we are proposing will significantly increase the number of data scientists that Michigan Tech can offer to the workforce. Our M.S. program in Data Science 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 minimum requirements listed in recent data scientist job postings include "strong communication and collaboration skills" (Groupon), "ability to communicate complex quantitative analysis in a clear, precise, and actionable manner" (Quicken Loans), and "expected to communicate their conclusions clearly to a lay audience" (CIA). The M.S. degree is not intended to be a stepping-stone towards a Ph.D.; rather, it is a stand-alone degree designed to prepare students for careers in 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 MichiganTechstrategicplan[8] to "be a leader in creating solutions for society's challenges through education and interdisciplinary endeavors that advance sustainable economic prosperity…"

---

[4] Data Scientists: Meet Big Data's Top Guns
http://www.informationweek.com/big-data/news/big-data-analytics/240006580/data-scientists-meet-big-da tas-top-guns,
8/21/12
[5] Data scientist: The hot new gig in tech
http://tech.fortune.cnn.com/2011/09/06/data-scientist-the-hot-new-gig-in-tech/ , 9/5/2011
[6] Data Science Revealed: A Data-Driven Glimpse into the Burgeoning New Field
http://www.emc.com/collateral/about/news/emc-data-science-study-wp.pdf, 2011
[7] Big data: The next frontier for innovation, competition, and productivity
http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
[8] STRATEGIC PLAN https://www.banweb.mtu.edu/pls/owa/strategic_plan2.p_display

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 master's 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 (2011), DePaul's M.S. in Predictive Analytics (2010), University of San Francisco's M.S. in Analytics (2012), LSU's M.S. in Analytics (2011), Rutgers's Professional Science Master's (PSM) of Business and Science in Analytics (2012), and NCSU's M.S. in Analytics (also a PSM program)(2007).

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.

4. Detailed Description of Master of Science in Data Science

i. Title:

Master of Science 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 in disciplines of science and engineering. The main threads of analytic techniques, programming practice, domain knowledge, business acumen, and communication skills are intertwined in this program.

iii. Relation to Professional Science Master's:

The M.S. degree is expected to meet the needs of students and to adhere to the requirements of the ProfessionalScienceMaster's[9] (PSM) programs. Students benefit from a PSM degree because it prepares them for careers in science and engineering that are highly sought after in industry, government, and nonprofit organizations, where workforce needs in data science are increasing. PSM graduates get advanced training in science and

[9] http://www.sciencemasters.com/

engineering without having to obtain a Ph.D., while simultaneously developing highly-valued business skills without having to obtain an MBA. The curricula for PSMs are based on "science-plus,[10]" where rigorous study in engineering, science, or mathematics is combined with skills-based coursework in management, policy, or law. In addition, PSM programs emphasize writing and communication skills and teamwork experience, with most requiring a "real-world" internship in an industry or public sector enterprise.

To comply with the PSM requirements, the M.S. program is grounded in science, technology, engineering, mathematics, computer science, and computing. It is designed to prepare students for a variety of careers that will fill the skill shortage in data science in industry, business, government, and non-profit organizations. This program prepares graduates for high-level careers in data science by combining advanced training in data analytics with an appropriate component of professional skills. In addition to the course work in data analytics and data management, the M.S. program will emphasize skill areas such as written and oral communication, ethics, management, policy, entrepreneurship, and leadership. These skills and experiences will be integrated as components of each of the core courses. The program will include an employer-sponsored project that will be incorporated in Data Science course "Introduction to Data Science UN5550.

Entry into this program assumes basic knowledge in statistical and mathematical techniques, programming, and communications, obtained through a degree in business, science and engineering disciplines. An entrance assessment exam may be used to test students competency and skills in their background knowledge. The purpose of the assessment exam is to determine student eligibility to the program and also to help guide students to courses to shore up their foundational knowledge (e.g., programming skills, statistical knowledge, etc.)

iv. Credits:

The degree will be offered as a 30-credit coursework-only M.S. program that meets the degree requirements of the Graduate School[11]. This includes the student completing training in the responsible conduct in research (RCR)[12].

v. Course work:

The M.S. in Data Science requires 12 credits of required core courses and a minimum of six credits of approved Data Science electives. The remaining required credits can include up to maximum of six credits of approved foundational courses at the 3000-4000 level (Appendix I), plus domain specific courses (Appendix II).

[10] "Science Plus" Curricula, http://www.sciencemasters.com/Default.aspx?tabid=83
[11] http://www.mtu.edu/gradschool/administration/academics/requirements/ms/
[12] http://www.mtu.edu/gradschool/administration/academics/resources/rcr/

Program admission skills

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. These skills may have been obtained through formal academic qualifications, work experience, or a

combination. 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. This may impact the number of credits required for a student; e.g., if a number of foundational courses are needed they may exceed 30 credits to complete the M.S.

Students will be allowed to apply up to six credits of 3000-4000 level foundational skills courses (Appendix I) toward the M.S. Data Science degree. Additional courses (more than 6 credits) of foundational skills courses cannot be applied to the degree even though additional courses may be required for students to master necessary skills. Courses at the 1000 or 2000 level cannot be applied toward any graduate degree at Michigan Tech but may be necessary for students to take if they are lacking in a key skill.

Each student's letter of offer of admission will clearly articulate the expectations for incoming students in the area of foundational skills in computer programming, statistical analysis, information systems and databases. Students will be encouraged to develop their foundational skills before coming to Michigan Tech to start the M.S. program in Data Science. They will be also advised of the availability of the low-level courses (1000 and 2000 level) and medium-level courses (3000 level) that they can take at Michigan Tech to prepare for the M.S. program in Data Science[13]. After 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.

---

[13] Courses offered during the summer are identified in Appendix I.

Coursework summary

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

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

- UN 5550 - Introduction to Data Science (Fall, 3 credits)[14]
- MA 4790 - Predictive Modeling (Fall, 3 credits)
- CS 4821 / MA 4795 - Data Mining (Spring, 3 credits)
- BA 5200 - Information Systems Management and Data Analytics (Fall, 3 credits)[15]

Foundational skills courses for M.S. Data Science (maximum of 6 credits at the 3000-4000 level):

A maximum of six credit hours of foundational skills courses at the 3000-4000 level may be applied to the M.S. These courses will build skills necessary for successful completion of the M.S. Data Science. See Appendix I for a list of approved foundational skills courses. Note, some students will not need to

take these foundational courses, and will instead use the domain electives to reach the credit requirements of the M.S. program.

[14] New course is designed for Fall 2014; submitted to curriculum proposal (binder) Fall 2013. This course will be administered by the Office of the Dean of the Graduate School, and executed by Graduate Program Executive Committee. It is listed as a UN course because it will involve faculty from multiple
department. It is the responsibility of the Graduate Program Director and the Graduate Program Executive Committee, with assistance provided by the Graduate School office to design the curriculum for this course, organize the weekly seminar lectures to be delivered by the Data Science faculty across the campus. The Graduate Program Director will work with the Chairs of the units involved to negotiate the faculty load and other resources.

[15] 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 6 credits):

At least 6 credits for the M.S. must be taken from the approved 3-credit Data Science elective courses below:

- CS 5841 / EE 5841 - Machine Learning (Spring, 3 credits)[16]
- CS 5491 - Cloud Computing (Spring, 3 credits)[17]
- CS 5471 – Advanced Topics in Computer Security (3 credits)[18]
- MA 5781 - Time Series Analysis and Forecasting (Spring, 3 credits)[19]
- BA 5740 - Managing Innovation & Technology (On Demand, 3 credits)
- PSY 5210 - Advanced Statistical Analysis and Design I (On Demand, 4 credits)
- FW 5083 - Bioinformatics Programming and Skills (Fall, 3 credits)[20]
- PH 4395 - Computer Simulation in Physics (on Demand, 3 credits)

Domain specific Data Science courses for M.S. program (maximum of 12 credits):

To complete the M.S. program in Data Science, the students must complete the remaining of the required 30 credits through completion of approved domain-specific Data Science courses (see Appendix II). Students may choose domain-specific courses from one or more domains. Each student will consult with her/his advisor in order to determine the appropriate mix of elective courses and domain-specific courses, given the student's background, interests, and career aspirations.

Three sample programs in domains of Computer Science/Computer Engineering, Management Information Systems, and Forestry are provided in Appendix IV.

[16] New course designed for Spring 2015; submitted to curriculum proposal (binder) Fall 2013

[17] New course designed for Spring 2015; submitted to curriculum proposal (binder) Fall 2013

[18] New course designed for Spring 2015; submitted to curriculum proposal (binder) Fall 2013

[19] 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.

[20] 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. Integration of essential business acumen, communication and teamwork skills in Data Science program:

The core and some of the approved elective courses are designed to integrate the skills in written and oral communication, ethics, management, policy, entrepreneurship, and leadership.

Core course UN 5500 requires presentation on a topic and written reports on a project from students. In this course students will apply ethical principles to the management
of confidential and sensitive data. They will analyze a business case study to
determine the appropriate level of information security associated with corporate data. Students will learn management tools and techniques appropriate for the role of data scientist. Guest speakers will provide personal experiences and management challenges encountered in their careers. Government regulations and corporate policies will be evaluated based on their relevance to the course content (e.g. Sarbanes-Oxley Act 2002, Gramm-Leach-Bliley Act, etc.). Guest speakers will deliver lectures on entrepreneurship opportunities in Data Science. Data scientists depend upon leadership skills to manage and disseminate predictive analytics results. Guest speakers will provide personal recollections on their leadership styles, and business
cases will be evaluated with respect to leadership techniques. Further, the program will include an employer-sponsored project that will be incorporated into this course.

Core course BA 5200 assignments require students to deliver oral presentations on hot topics in the Information Systems /Information Technology field. They are also required
to present their final team report at the end of the semester. There is one lecture specifically on ethical principles; the ethical context is discussed throughout the course. One writing assignment requires graduate students to apply ethics to an ethical dilemma involving an Information Systems breach (business case analysis). This class covers data and Information Technology governance within a business context, focusing on corporate policies and procedures. Entrepreneurship is covered only briefly in this course, as it relates to data management and decision making. Leadership and management theories are discussed at length throughout the course,
from different perspectives (e.g. Project Management, Chief Information Officer or Chief Technical Officer, etc.)

Core course MA 4790 and approved elective MA 5781 would emphasis written and oral communications by requiring students to complete a computational group project with
a written and oral reports. Management and leadership are dealt with through group projects that involve managerial and leadership skills. The ethical consequence of predictive analytics that applies statistical techniques to the analysis of large data sets, to produce actionable intelligence is presented in the course.

Course CS 4821 requires each student to give one to two short in-class presentations on recent news and examples in the field of Data Mining. The course has the students learn in class about issues of data privacy, read several academic and popular press articles on the subject, and participate in several discussion posts reflecting on these topics. As part of their final project, students will be

required to perform a short literature review on the entrepreneurship of their elected topic. Other requirements in reference to the final project include oral presentations to the class and a final written report (in the style of a conference paper). The final report requires a draft submission that is peer-reviewed by other students in the class.

Approved elective CS 5491 requires each student to make three to four written and in-class presentations on recent topics relevant to the course. It also includes ethical challenges of data in the cloud.

Approved elective BA 5740 assesses written communication via weekly reflection papers on the assigned readings. Oral communication is assessed via two group presentations in class. Ethics are introduced as part of the discussion about the consequences of the failure of firms, barriers to new entry, and innovation strategies. Management is discussed in the context of organizing teams and creating organizational structures conducive to innovation. Policy (i.e., strategy) is a core part of the course as we are interested in business models that are disruptive to incumbents. Entrepreneurship is covered to the extent that disruptive innovations are typically brought to market by new entrants; we also cover the managerial decision process for deciding which complementary assets to internalize or outsource. Finally, leadership,
or strategic leadership, is viewed as an important behavior in ambidextrous organizations allowing them to explore and exploit opportunities simultaneously.

vii. Online delivery:

Our goal is to have all the core data science courses and most approved data science courses offered online by 2016. This would allow off-campus students to complete 12-18 credits of the M.S in Data Sciences online[21]. 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.

viii. Description of content covered in new or revised Data Science courses:

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

---

[21] This would also allow off-campus students to fully complete the Graduate Certificate in Data Sciences with online offerings.

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.[22]

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.

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);

[22] 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

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.

SS 5005 - Introduction to Computational Social Science (new) (3 credits) (See Appendix II)

An introduction to computational methods for the social sciences. The course provides an introduction to complexity theory and Agent-Based Modeling. Students will apply what they have learned in this course to develop a pilot simulation to understand any social phenomena of their choosing.

SAT 5600 - Web Application Development (new) (3 credits) (See Appendix II)

This course provides an introduction to the building and administration of web applications. Topics covered include Apache web server, Tomcat application server, HTML, cascading style sheets,JavaScript, JQuery, server side includes, server side application development, web services, Secure Sockets Layer/Transport Layer Security and authentication/ authorization.

SAT 5002 - Application Programming Introduction (new) (3 credits) (See Appendix II)

This course provides an introduction to application programming. It develops problem solving skills through the application of a commonly used high-level programming language. Topics include the nature of the programming environment, fundamentals of programming languages (e.g., programming constructs, data management, manipulation of simple data structures), structured programming concepts, object oriented programming concepts, desirable programming practices and design, debugging and testing techniques. Students will use the Java programming language to test programming concepts and to develop application programs.

5. Estimated Costs For Financial Evaluation

While hard to predict the exact figures, it is expected the program will have an initial enrollment of five in Fall 2014, with an increase to eight in Fall 2015. After accreditation as a PSM program it is

expected that we will have an steady enrollment of 15 to 20 within five years.

To arrive at this estimation a combination of factors were considered; enrollment in other institutions in Data Science and PSM programs, enrollment in other M.S. programs in Michigan Tech, frequent recent requests from the industry for skills analytics (e.g., Kimberly-Clark), feedback from ECE EAC members, etc.

It is envisioned that the enrollment would primarily come from students who would otherwise

not have come to MTU. However, feedback from several units across the University indicates to the likely popularity of the courses that are designed for Data Science in other existing programs.

The initial start up cost of this program is modest. The majority of courses in the Data Science program are based on existing courses in Mathematical Sciences, Computer Science, and the School of Business and Economics. However, the quality of the program is directly dependant on the ability of the core faculty to develop and continually improve the core and approved elective Data Science courses. To ensure long-term viability of the Data Science program at Michigan Tech, therefore, requires sufficient allocation of resources to support the core Data Science curriculum. The program requires additional resources for the following:

- One new faculty line to be used to attract a strong candidate with multi-disciplinary data science expertise who will be dedicated to the Data Science program.[23] The new faculty will be primarily responsible for teaching of the core and Data Science courses in Data Science program. Michigan Tech should consider this faculty line as a strategic investment in an important area for the university. A faculty search procedure will be proposed by the Data Science Executive Committee (Sec. 7), approved by the the Dean of the Graduate School, and recommended to the appropriate Department Chairs, Deans, and Provost. A joint appointment would be anticipated. The anticipated hire will also help Michigan Tech to enhance research expertise in the area of Data Science. We are in a position to start the program with the existing pool of faculty resources in the four units involved in the offering of core and approved elective Data Science courses, through the temporary diversion of resources. However, the program requires the hiring of one new faculty after the first year.

- Development of core course—UN5550 Introduction to Data Science. This course will be non-departmental; select faculty from Mathematical Sciences, Computer Science, and School of Business & Economics will have a leading role in the administration of this new course. Modest cost may be involved in providing faculty adequate preparation time. There is also need for resources to cover the cost associated with the external guest lectures.

- Development of approved Data Science courses, Machine Learning (CS 5841/EE 5491) and Cloud Computing (CS 5491). These courses have also been planned for inclusion in the Electrical & Computer Engineering and Computer Science M.S. programs. The program also requires revision of BA 5670 into IS Management and Business Analytics, in the School of Business and Economics, to make it suitable to

---

[23] The Provost and Deans have committed the following resources that align with the Data Science Program. In the Department of Mathematical Sciences, there are currently searches for a senior and junior faculty position in Statistics; they will contribute to Data Science. In addition the provost supports a new lecturer line in Statistics starting Fall 2014. There is also a university commitment to fill a junior faculty line by Fall 2015 in one of the gap areas identified in the Computing, Information and Automation area. Data Science, next to Cyber Security, is one of them.

the Data Science program. There will be also a new domain specific course in the Social Sciences—Introduction to Computational Social Science. Modest cost may be involved in faculty preparation time for the development of these new courses.

- Administration of the program by the Graduate School, extension to online delivery by Fall 2016, and PSM affiliation will incur some additional cost.[24] We envision a ¼ to ½ line of administrative support when the program develops into its capacity of steady enrollment.

- Annual PSM membership, affiliation, and reaffiliation review fee of $1,500.

- Three graduate teaching assistant (GTA) lines are needed initially to help with the laboratory development and maintenance of the hardware and software tools necessary for the program.[25] Specifically these GTAs will be used to assist with the instruction of core and approved elective Data Science courses listed in Section 4. It is envisioned that the students supported by these three GTA lines will be earning Ph.D. degrees in disciplines related to the Data Science program. The purpose of the GTAs is to support the offering of cutting-edge courses in order to allow affiliated
faculty to maintain their active research programs. These lines will also help us to build our pool of research expertise in the area across the University. After the first three years of the project, the number of GTA lines allocated in support of the program may be reduced to one if circumstances allows it.

- Student support through theGraduateTuitionGrant(GTG)[26] which is designed to assist high-performing domestic students with unmet financial need. This program is already in existence and we expect to enhance the size of the pool of funding available to students by seeking external donations. We view fellowship and scholarship support for this program as important resources for getting this program started. There is often a time-lag between the start of a new program and being able to obtain external
support for students in the program; hence, this initial investment in student support will fill that gap. We also plan to seek external funding for the following:

    - A fellowship for an outstanding Data Sciences M.S. candidate, to be named in honor of Professor Thomas Drummer who was instrumental in developing the Data Science program.

    - Several smaller scholarships for excellent students pursuing an M.S. in Data Sciences.

- The Office of Information Technology at Michigan Tech is fully supportive of this program. It has already installed all the tools and computing infrastructure that is needed for the Data Science program.

---

[24] The Provost and Deans have committed staff support for the purposes of assisting prospective applicants for Data Science and the other interdisciplinary and non-departmental programs housed within the graduate school.

[25] The Provost and Deans have committed to one GTA from existing resources.

[26] Graduate Tuition Grant, http://www.mtu.edu/gradschool/admissions/financial/tuition-grant/

The resources of the Data Science program and the host departments will require continuous evaluation to ensure that the needs of each are being met in the future.

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 core and approved elective courses.

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 overall responsibility and final oversight for the program. The program will have the following management structure.

- **Graduate Program Director**: The Director is appointed by the Dean of the Graduate School for a period of three years. The Dean of the Graduate School will seek nominations for the

Graduate Program Director position from the Graduate Program Executive Committee. The Graduate Program Director will report to the Dean of the Graduate School and the Graduate Program Executive Committee. The Graduate Program Director will serve as the interim advisor for all incoming students until such time that each student identifies or is assigned a permanent advisor. The Graduate Program Director will meet with the Dean of the Graduate School and the Graduate Program Directors for other non-departmental and interdisciplinary graduate programs on a regular basis.

- **Graduate Program Executive Committee**: This committee is drawn from and elected by the membership of the Graduate Program Faculty (see next item). The committee will consist of three to five members that are representative of the diversity of programs and areas of interest of the Graduate Program Faculty. Members will serve for staggered five year terms. The Program Director serves as an ex-officio member of the Graduate Program Executive Committee. A staff member from the University IT Services will serve in a non-voting advisory capacity on the Graduate
Program Executive Committee. This group will work with the Program Director to make day-to-day decisions regarding the program. This group will identify potential members of the Graduate Program Faculty and External Advisory Board (see below) and
conduct voting in order to determine the membership of those bodies. The Graduate Program Executive Committee will provide leadership for the program and will organize and contribute to meetings of the External Advisory Board. Members of the Graduate Program Executive Committee will annually review the membership of the Graduate

  Program Faculty and External Advisory Board and recommendations for additions or removals will be made to the Dean of the Graduate School.

- **Graduate Program Faculty**: The faculty for this body is drawn from a wide range of units across the Michigan Tech campus community are affiliated with the Data

  Science program. These faculty members will have adjunct faculty[27] status in the Data Sciences program and will be eligible to advise students who are seeking degrees in Data Sciences. Appointments to the Graduate Program Faculty will be made for terms of three-years duration, with the possibility of reappointment for multiple successive terms. Members of the Graduate Program Faculty will be expected to participate in Graduate Program meetings and events. Graduate Program Faculty may be elected to serve as a member of the Graduate Program Executive Committee, and upon this election may be nominated to serve as the Graduate Program Director. Graduate Program Faculty are encouraged to advise the Graduate Program Director on the direction of the Data Science program at Michigan Tech, the development of
resources, and creation of opportunities for growth. Additionally, the Graduate
Program Faculty are encouraged to actively network with industry experts.

- **External Advisory Board**: This board is drawn from a key pool of experts from industry and academia operating in the forefront of Big Data science. This board will help ensure that the Michigan Tech Data Science program is abreast of industry needs.This board will act as an advocate for the program through its wide-reaching network. An External Advisory Board is a required component of all PSM (Professional Science Master's) programs. Members of the External Advisory Board will serve staggered three year terms. Members will be allowed to serve multiple terms.

- **Tentative membership**: Appendix III list the tentative membership of three functional bodies for the 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 a maximum of six credit hours of foundational skills courses at the 3000-4000 level may be applied to the M.S. These courses will build skills necessary for successful completion of the M.S. Data Science. 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 M.S. in Data Science.

Note that 2000 level courses listed here cannot be counted towards the requirement for M.S. 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) [28]
- SAT 3210 - DB Management (Credits: 3)[29]
- SAT 4600 - Web Application Development (Credits: 3)[30]

Appendix II: Domain Specific Data Science Courses

Mathematics Courses (Credits: 3)

- MA 4710 - Regression Analysis
- MA 4720 - Design and Analysis of Experiments
- MA 4330 - Linear Algebra
- MA 5201 - Combinatorial Algorithms
- MA 5221 - Graph Theory
- MA 5401 - Real Analysis
- MA 5627 - Numerical Linear Algebra
- MA 5630 - Numerical Optimization
- MA 5701 - Statistical Methods
- MA 5741 - Multivariate Statistical Methods
- MA 5750 - Statistical Genetics
- MA 5761 - Computational Statistics
- MA 5791 - Categorical Data Analysis

Computer Science Courses (Credits: 3)

- CS 4425 - Data Management System Design
- CS 4471 - Computer Security
- CS 5321 - Advanced Algorithms
- CS 5331 - Parallel Algorithm
- CS 5441 - Distributed System
- CS/EE 5496 - GPU and Multi-core Programming
- CS 5631 - Data Visualizations
- CS 5760 - HCI Usability Testing
- CS 5811 - Advanced Artificial Intelligence
- CS/EE 5821 - Computational Intelligence

Department Computer Science may also consider developing new courses when appropriate or necessary in visual analytics, mobile applications and graduate software engineering service course.

School of Technology Courses (Credits: 3)

- SAT 5001 - Introduction to Medical Informatics
- SAT 5002 - Application Programming Introduction[31]
- SAT 5121 - Introduction to Medical Sciences, Human Pathophysiology, Healthcare
- SAT 5141 - Clinical Decision Support and Improving Healthcare
- SAT 5161 - Data Warehousing and Business Intelligence
- SAT 5241 - Designing Security Systems
- SAT 5600 - Web Application Development[32]

- SU 5010 - Geospatial Concepts, Technologies and Data
- SU 5045 - Geospatial Data Fusion

Electrical and Computer Engineering Courses (Credits: 3)

- CS/EE 5496 - GPU and Multi-core Programming
- EE 5500 - Probability and Stochastic Processes
- EE 5521 - Detection & Estimation Theory
- EE 5726 - Embedded Sensor Networks
- CS/EE 5821 - Computational Intelligence

Civil and Environmental Engineering Courses (Credits: 3)

- SSE 3200 - Web Based Services
- CE/SSE 4750 - Risk Analysis
- CE/SSE 4760 - Optimization and Decision-making
- CE/SSE 5710 - Modeling and Simulation Applications for Decision-Making in Complex Dynamic Systems.
- CE 5740 - System Identification

School of Business and Economics Courses (Credits: 3)

- MIS 3100 - Business Database Management
- MIS 3400 - Business Intelligence
- EC 4200 - Econometrics
- BA 5610 - Business Process Management

[31] New 3-credit course designed for Fall 2014; submitted to curriculum proposal (binder) Fall 2013
[32] New 3-credit course designed for Spring 2015; submitted to curriculum proposal (binder) Fall 2013

- BA 5800 - Marketing, Technology, and Globalization

Geological and Mining Engineering and Sciences Courses (Credits: 3)

- GE 5150 - Advanced Natural Hazards
- GE 5195 - Volcano Seismology
- GE 5250 - Advanced Computational Geosciences
- GE 5600 - Advanced Reflection Seismology
- GE 5670 - Aquatic Remote Sensing
- GE 5870 - Geostatistics & Data Analysis

School of Forestry Courses

- FW 5084 - Data Analysis and Graphics Using R (Credits:2)
- FW 5089 - Tools of Bioinformatics (Credits:4)
- FW 5411 - Applied Regression Analysis (Credits:3)

- FW 5412 - Regression with the R Environment for Statistical Computing (Credits:1)
- FW 5540 - Advanced Terrestrial Remote Sensing (Credits:4)
- FW 5550 - Geographic Information Systems for Resource Management (Credits:4)
- FW 5555 - Advanced GIS Concepts and Analysis (Credits:3)
- FW 5556 - GIS Project Management (Credits:3)
- FW 5560 - Digital Image Processing: A Remote Sensing Perspective (Credits:4)

Social Science Courses

- SS 5005 - Introduction to Computational Social Science (Credits: 3)[33]
- SS 5315 - Population and Environment (Credits: 3)

Department of Social Sciences may also consider developing new courses, when appropriate or necessary in computational social sciences with elements of social science theory, and land use modeling.

Cognitive and Learning Sciences Courses

- PSY 5220 - Advanced Statistical Analysis and Design II (Credits: 4)

Biological Sciences Courses

- BL 4470 - Analysis of Biological Data (Credits:3)

[33] New 3-credit course designed for Fall 2014; submitted to curriculum proposal (binder) Fall 2013

Biomedical Engineering Courses

- BE 5550 - Biostatistics for Health Research (Credits: up to 4)

Department of Biomedical Engineering may also consider developing new courses, when appropriate or necessary in big data applications to human health. This course may be developed as a BE, BL, or KIP course.

Chemical Sciences Courses (Credits: 3)

- CH 4610 - Introduction to Polymer Science
- CH 5410 - Advanced Organic Chemistry: Reaction Mechanisms
- CH 5420 - Advanced Organic Chemistry: Synthesis
- CH 5509 - Transport and Transformation of Organic Pollutants
- CH 5515 - Atmospheric Chemistry
- CH 5516 - Aerosol and cloud chemistry
- CH 5560 - Computational Chemistry

Department of Chemistry may consider developing new courses when appropriate or necessary in bio-spectroscopy and cheminformatics

Physical Sciences Courses

- PH 4390 - Computational Methods in Physics (Credits: 2)

Appendix III: Tentative Membership of Data Science Bodies

Graduate Program Faculty Membership

- Laura Brown (Computer Science)
- Mari W. Buche (School of Business & Economics)
- Jason Carter (Kinesiology)
- Sarah Green (Chemistry)
- Timothy Havens (Electrical and Computer Engineering/Computer Science)
- Guy Hembroff (School of Technology)
- Chandrashekhar Joshi (Biological Sciences)
- Sarah Lucchesi (Library)
- Robert Nemiroff (Physics)
- Saeid Nooshabadi (Electrical & Computer Engineering/Computer Science)
- Thomas Oommen (Geological & Mining Engineering & Sciences)
- Mark Rouleau (Social Sciences)
- Gowtham S (Information Technology Services)
- Ching-Kuang Shene (Computer Science)
- Allan Struthers (Mathematics)
- Raymond Swartz (Civil and Environmental Engineering)
- Hairong Wei (Forestry & Bioinformatics)

Graduate Program Executive Committee Membership

- Laura Brown (Computer Science)
- Mari W. Buche (School of Business & Economics)
- Timothy Havens (Electrical & Computer Engineering/Computer Science)
- Saeid Nooshabadi (Electrical & Computer Engineering/Computer Science)
- Gowtham S (Information Technology Services)
- Allan Struthers (Mathematics)

External Advisory Board Membership

- David Barnes (Program Director, Strategy and Emerging Internet Technologies, IBM)
- Tom Grebinski (Founder, Yotta Data Sciences)
- Lonne Jaffe (CEO of Syncsort)
- Jill Recla (Bioinformatics Analyst at The Jackson Laboratory)

- John Soyring (Soyring Consulting Services)

- John Wallin (Professor of Physics and Astronomy, and Director of the Computational Sciences Program at Middle Tennessee State University)

Appendix IV: Sample Data Science Schedules

Sample Schedule for student with background and domain focus in computational methods (computer science/computer engineering/etc.)

Students with this background are expected to have the foundational skills in computer programming, statistical analysis, information systems and databases. Therefore, no foundational classes appear in the schedule.

Legend

Core Courses	Foundational Skills
Approved Electives	Domain Electives

Year 1

Fall	Spring
UN 5550 – Intro to Data Sci	CS 4821 – Data Mining
MA 4790 – Predictive Modeling	Domain Elective – 3 cr. *
BA 5200 – Info. Sys. Mgmt. & Bus. Analytics	CS 5491 – Cloud Computing

18 credits

Year 2

Fall	Spring
Domain Elective – 3 cr. *	CS 5841 – Machine Learning
Domain Elective – 3 cr. *	Domain Elective – 3 cr. *
Domain Elective – 3 cr. *	Domain Elective – 3 cr. *

18 credits

* Note, the domain electives are not restricted by background, that is, a student with a BS in computer science can take electives in any other department as long as any pre-requisites are met.

Sample Schedule in for student background and domain focus in Information Systems (MIS)

In this sample schedule the prerequisites necessary for some of the MA and CS classes are not met by the business degree requirements. Therefore, prerequisites foundational courses are incorporated into the sample schedule.

Legend

Core Courses	Foundational Skills
Approved Electives	Domain Electives

Year 1

Fall	Spring
UN 5550 – Intro to Data Sci	MIS 3100 – Business Database Mgmt.
MA 3740 – Statistical Prog. and Analysis	Domain Elective – 3 cr.
BA 5200 – Info. Sys. Mgmt. & Bus. Analytics	Domain Elective – 3 cr.

18 credits

Year 2

Fall	Spring
MA 4790 – Predictive Modeling	CS 4821 – Data Mining
Domain Elective – 3 cr.	MA 5780 – Time Series Analysis & Forecasting

Domain Elective – 3 cr.	BA 5740 - Managing Innovation & Technology

18 credits

Sample Schedule in for student background and domain focus in forestry

In this sample schedule the prerequisites necessary for some of the MA and CS classes are not met by the forestry degree requirements (programming and database skills).   Therefore, prerequisites foundational courses are incorporated into the sample schedule.

Legend

Core Courses	Foundational Skills
Approved Electives	Domain Electives

Year 1

Fall	Spring
UN 5550 – Intro to Data Sci	MIS 3100 – Business Database Mgmt.
MA 3740 – Statistical Prog. and Analysis	FW 5083 – Bioinformatics Prog. and Skills
Domain Elective – 3 cr.	Domain Elective

18 credits

Year 2
Approved by Senate: 26 March 2014

Fall	Spring
MA 4790 – Predictive Modeling	CS 4821 – Data Mining
BA 5200 – Info. Sys. Mgmt. & Bus. Analytics	Domain Elective FW 5411 – Applied Regression Analysis*
Domain Elective FW 5412 – Regression with R for Stat. comp*	MA 5780 – Time Series Analysis & Forecasting

14-18 credits

* Note, the domain electives are added as suggestions and are subject to their availability and interest of the student. Also, the domain electives are not restricted by background, that is, a student with a BS in forestry can take electives in any other department as long as any
pre-requisites are met.

Introduced to Senate: 05 March 2014
Approved by Senate: 26 March 2014
Approved by Administration: 03 April 2014
Approved by BOC: 02 May 2014
Approved by State: 05 June 2014