Major Program Modification

New Computer Science Specialization – Data Science

Prepared by Ken Pu and Faisal Qureshi

1. Introduction

According to a new report by McKinsey Global Institute, we are living through "Digital Globalization," an era where "soaring flows of data and information … generate more economic value than the global goods trade."[1] There has been an explosive growth in the amount of data that has been generated and made available to companies, governments and individuals. The term Big Data has been coined to address the incredible value of intelligent analysis of large volumes of data to support crucial business and economical decisions. We urgently need individuals capable of turning vast quantities of data, in other words Big Data, into insight. Individuals with this ability, often called *data scientists*, are highly sought after in a variety of industries, from healthcare to finance to business consultancy. It is no surprise that Harvard Business Review calls data scientist, "the sexiest job of the 21st century."[2]

Data scientists need strong mathematical, statistical and programming skills to grapple with massive amount of structured and unstructured data with the aim to drive value from this data. These individuals are able to 1) extract data from diverse sources, 2) organize it, 3) construct predictive, prescriptive, and prognostic models from this data, 4) examine and analyze data using these models, 5) develop and implement new algorithms to manage data flow, 6) explore trends and infer hidden relationships through data inspection and data visualization, and 7) communicate findings to the management.

The undergraduate Bachelor of Science Computer Science program already offers a set of courses that are highly relevant data scientists. These courses, for example, deal with data capture, archival, interaction, visualization and analysis. These courses, we believe, provide excellent opportunities for data scientists' training. *Consequently, we propose to develop and offer Data Science specialization as part of the existing undergraduate Bachelor of Science Computer Science program.*

Data Science is inherently multi-disciplinary and straddles computer science, mathematics, statistics, plus other more domain specific fields, such as physics, chemistry and biology. The Faculty of Science with its breadth of scientific disciplines offers additional advantages for Data Science specialization. Our students will not only develop fundamental computational skills for analyzing large volumes of data, should they chose to, the students will also have the unique opportunity to gain domain specific knowledge by focusing on one or more disciplines available within the faculty.

---

[1] Manyika et al., Digital globalization: The new era of global flows, McKinsey Global Institute, Feb 2016.

[2] Daveport and Patil, Data Scientist: The sexiest job of the 21st century, Harvard Business Review, Oct 2012.

2. Degree Requirements

a. Program Learning Outcomes

Graduates of the proposed Data Science specialization are expected to have a solid understanding of Computer Science. Additionally, these students will focus on developing knowledge, skills, and tools that will set them for success in the highly lucrative and desirable Data Science job market. Data Science specialization students will develop problem solving, critical, analytical, computational, and communication skills that will enable them to turn vast quantities of data into insight. Specific learning outcomes are:

- Develop a solid understanding of the theoretical foundation of data science;
- Achieve a strong command over programming and computational tools to effectively handle data with volume and complexity;
- Gain working knowledge over a range of algorithms, methods and systems for performing data analysis on real-life data sets;
- Attain a good understanding of the applications of data science across a variety of domains; and
- Acquire critical and problem solving skills to extract knowledge and understanding from large volumes data.

b. Admission Requirements

No changes to the admission requirements will be made to accommodate the new proposed Data Science specialization.

c. Program Structure

Table 1 summarizes the program map for Data Science specialization. Courses in blue are required courses for students in Data Science specialization. These courses are available to students in Computer Science Comprehensive program and Computer Science Digital Media specialization; however, these courses are not required for these students.

CSCI 1030U Introduction to CS	CSCI 1060U Programming Workshop I	MATH 1000 Calculus	PHY 1010U Physics	Elective
CSCI 2050U Computer Architecture	CSCI 1061U Programming Workshop II	MATH 1020U Calculus II	PHY 1020U Physics II	Elective
CSCI 2000U Scientific Data Analysis	CSCI 2010U Principles of CS	CSCI 2110U Discrete Structures in CS	STAT 2010U Statistics and Probability	Elective
CSCI 2020U Software System Dev & Int.	CSCI 2040U Software Design & Analysis I	MATH 2050U Linear Algebra	CSCI 2072U Computational Science I	Elective
CSCI 3230U Web Application Development	CSCI 3030U Database	CSCI 3070U Algorithms	CSCI 4040U Ethics, Law and Social Impact	Elective
CSCI 3055U Programming Languages	CSCI 3060U Software Quality Assurance	CSCI 3090U Comp. Graphics & Visualization	CSCI 4210U: Info Visualization	Elective
CSCI 4410U Thesis I	CSCI 3010U Simulation and Modeling	CSCI XXXX: Machine Learning or CSCI 4610: Artificial Intelligence	CS Elective	Elective
CSCI 4420U Thesis II	CSCI 4020U Compilers	CSCI 4030U Big Data	CSCI 4220U Computer Vision	Elective

Table 1 Computer Science Data Science Specialization program map. Courses in blue indicate changes from the CS comprehensive program.

Proposed Calendar Copy

14.8.4.3 Computer Science – Data Science specialization
YEAR 1 – Regular program and Co-operative Education program

Semester 1 (15 credit hours)
CSCI 1030U Introduction to Computer Science
CSCI 1060U Programming Workshop I
MATH 1000U Introductory Calculus[+] or MATH 1010U Calculus I[+]
PHY 1010U Physics I[+] or PHY 1030U Introductory Physics[+]
Elective**

Semester 2 (15 credit hours)
CSCI 1061U Programming Workshop II
CSCI 2050U Computer Architecture
MATH 1020U Calculus II
PHY 1020U Physics II
Elective**

[+]All students who have completed Grade 12 Advanced Functions (MHF4U) and Calculus and Vectors (MCV4U) should take MATH 1010U and PHY 1010U. Students without one of these high school courses or equivalent are directed to take MATH 1000U and PHY 1030U.

YEAR 2 – Regular program	**YEAR 2 – Co-operative Education program**
Semester 1 (15 credit hours)	**Semester 1 (15 credit hours)**
CSCI 2000U Scientific Data Analysis	CSCI 2000U Scientific Data Analysis
CSCI 2010U Principles of Computer Science	CSCI 2010U Principles of Computer Science
CSCI 2110U Discrete Structures in Computer Science	CSCI 2110U Discrete Structures in Computer Science
STAT 2010U Statistics and Probability for Physical Science	STAT 2010U Statistics and Probability for Physical Science
Elective**	Elective**
Semester 2 (15 credit hours)	**Semester 2 (15 credit hours)**
CSCI 2020U Software Systems Development and Integration	CSCI 2020U Software Systems Development and Integration
CSCI 2160U Digital Media	CSCI 2160U Digital Media
CSCI 2040U Software Design and Analysis	CSCI 2040U Software Design and Analysis
CSCI 2072U Computational Science I	CSCI 2072U Computational Science I
MATH 2050U Linear Algebra	MATH 2050U Linear Algebra
	Semester 3
	SCCO 1000W Co-op Work Term I*

YEAR 3 – Regular program **Semester 1 (15 credit hours)** CSCI 3230U Web Application Development CSCI 3030U Database Systems and Concepts CSCI 3070U Analysis and Design of Algorithms CSCI 4040U Ethics, Law and the Social Impact of Computing Elective** **Semester 2 (15 credit hours)** CSCI 3055U Programming Languages CSCI 3060U Software Quality Assurance CSCI 3090U Computer Graphics and Visualization CSCI 4210U Information Visualization Elective**	**YEAR 3 – Co-operative Education program** **Semester 1 (15 credit hours)** CSCI 3230U Web Application Development CSCI 3030U Database Systems and Concepts CSCI 3070U Analysis and Design of Algorithms CSCI 4040U Ethics, Law and the Social Impact of Computing Elective** **Semester 2** SCCO 2000W Co-op Work Term II* **Semester 3** SCCO 3000W Co-op Work Term III*
YEAR 4 – Regular program **Semester 1 (15 credit hours)** CSCI 3010U Simulation and Modeling CSCI 4610U Artificial Intelligence or CSCI XXXX Machine Learning CSCI 4410U Computer Science Thesis Project I*** or Senior Computer Science elective** Elective** CS Elective** **Semester 2 (15 credit hours)** CSCI 4020U Compilers CSCI 4030U Big Data CSCI 4220U Computer Vision CSCI 4420U Computer Science Thesis Project II*** or Senior Computer Science elective** Elective**	**YEAR 4 – Co-operative Education program** **Semester 1** SCCO 4000W Co-op Work Term IV* **Semester 2 (15 credit hours)** CSCI 3055U Programming Languages CSCI 3060U Software Quality Assurance CSCI 3090U Computer Graphics and Visualization CSCI 4210U Information Visualization Elective** **Semester 3** SCCO 5000W Co-op Work Term V*
	YEAR 5 – Co-operative Education program **Semester 1 (15 credit hours)** CSCI 3010U Simulation and Modeling CSCI 4610U Artificial Intelligence or CSCI XXXX Machine Learning CSCI 4410U Computer Science Thesis Project I*** or Senior Computer Science elective** Elective** CS Elective**

	Semester 2 (15 credit hours) CSCI 4020U Compilers CSCI 4030U Big Data CSCI 4220U Computer Vision CSCI 4210U Information Visualization CSCI 4420U Computer Science Thesis Project II*** or Senior Computer Science elective** Computer Science elective**

Notes:
No more than 42 credit hours may be taken at the first-year level.

*This course is graded on a pass/fail basis.

****Electives and breadth requirements**
Students must complete a total of 24 elective credit hours such that the following requirements are satisfied:

- 15 credit hours must be in courses offered by the Faculty of Science, of which at least 6 credit hours
  must be in Computer Science electives.
- 9 credit hours must be in courses from outside the Faculty of Science, among which at least 3 credit hours must be in business electives[++], and at least 3 credit hours in communications electives.[+++]

Computer Science electives for the Data Science specialization:
CSCI 3050U Computer Architecture II
CSCI 3220U Digital Media Production
CSCI 4120U Digital Evidence
CSCI 4100U Mobile Devices
CSCI 4130U Forensic Informatics
CSCI 4620U Human-Computer Interaction
CSCI 4630U High-Performance Computing
CSCI 4640U Distributed Computing
CSCI 4650U Elements of Theory of Computation
MATH 2015U Calculus III
MATH 3030U Introduction to Probability Theory*
MATH 3050U Mathematical Modelling**
STAT 3010U Biostatistics

* students are required to take MATH 2015U before they can take MATH 3030U
** students are required to take MATH 2015U and one of MATH 2060U or MATH 2860U before they can take MATH 3050U

[++]**Business electives:**
BUSI 1020U Business Communications
BUSI 1600U Management of the Enterprise
BUSI 1700U Introduction to Entrepreneurship

BUSI 2000U Collaborative Leadership

[+++]**Communication electives:**
COMM 1100U Introduction to Communication
COMM 1050U Technical Communication
COMM 1310U Fundamentals of Professional Writing
COMM 1320U Oral Communication and Public Speaking
COMM 1610U Interpersonal Communication

[***]**Thesis Project or senior Computer Science elective courses**
Students in clear academic standing who have completed 90 credit hours of their program and six third-year required courses may optionally apply to take a two course sequence consisting of CSCI 4410U and CSCI 4420U (Computer Science Thesis Project I and II). Students not accepted to take the thesis courses must complete two additional Computer Science electives instead. A student meeting the above requirements who does not take CSCI 4410U and CSCI 4420U may optionally apply to take CSCI 4430U (Directed Studies in Computer Science) as one of the required computer science electives. Opportunities for the Thesis Project and Directed Studies courses are limited; students must apply through Science Advising by March 30 following completion of the first three years of the program.

d. Program Content

Table 1 shows the program map for the proposed Data Science specialization.

3. Resource Requirements

a. Faculty Members

Core faculty associated with this specialization include all current Computer Science core faculty (tenure & tenure-track and teaching streams):

- Jeremy Bradbury, *Associate Professor*;
- Christopher Collins, *Assistant Professor and CRC in Linguistic Information Visualization*;
- Mark Green, *Professor*;
- Randy Fortier, *Lecturer*;
- Ken Pu, *Associate Professor*;
- Faisal Qureshi, *Associate Professor*; and
- Jaroslaw Szlichta, *Assistant Professor*.

No new core faculty hires are required.

b. Additional Academic and Non-Academic Human Resources

No new courses are being created as a part of the proposed specialization. The specialization rather restricts the courses that students can take in the 3rd and 4th years of their study (year 5 for Co-operative Education). We already offer these courses as (see Table 2 below):

- Computer Science electives for Computer Science Comprehensive program and Computer Science Digital Media specialization; and
- required courses for Computer Science Digital Media specialization.

Course	Current offering	Changes
CSCI 4030U Big Data Analytics	CS Elective	Required by the proposed specialization
CSCI 4210U Information Visualization	CS Elective ½ required by Digital Media Specialization	Required by the proposed specialization
CSCI 4220U Computer Vision	CS Elective ½ required by Digital Media Specialization	Required by the proposed specialization
CSCI 3010U Simulation and Modeling	CS Elective	Required by the proposed specialization

Table 2 Courses that we have offer as a part of the proposed Data Science specialization.

Summary

- Offering the proposed specialization simply limits our ability to offer "other" Computer Science electives without extra resources.
- We will not offer any new electives. We are required to offer 6 upper year Computer Science Electives every year (3 during the fall term and 3 during the winter term).
- We expect that the proposed Data Science specialization will have no significant resource impact, since we are using courses that we already offer as a part of Computer Science Comprehensive program and Computer Science Digital Media specialization.

d. Physical Resource Requirements

The proposed Data Science specialization has no extra physical resource requirements, since we don't plan to offer any extra elective Computer Science courses.

4. Business Plan

a. Statement of Funding Requirements

The proposed Data Science specialization has no extra funding requirements.

b. Statement of Resource Availability

The proposed specialized as outlined here does not require any new resources.

Faculty Council Approval: May 11, 2016

TEMPLATE 8-A

NEW COURSE TEMPLATE

For changes to existing courses see Course Change Template

Faculty: Science		
Course title: ***Machine Learning, Theory and Application***		
Course number: 4XXX	**Cross-listings:** CSCI	**____ Core _X_ Elective**
Credit weight: 3	**Contact hours:** __3__ Lecture __1.5__ Lab _____ Tutorial _____ Other	

CALENDAR DESCRIPTION

Machine learning is a branch of Computer Science that enables machines to identify patterns, make predictions and organize data by synthesizing models of the world through learning. In this course, we will cover the theory and application of machine learning. We will provide a survey of the fundamental building blocks of machine learning covering areas such as general probabilistic models and parameter estimation, regression models, statistical data analysis, neural networks and neural computation. We will place special emphasis on the application of the machine learning techniques in data representation, pattern recognition, classification and prediction. Students will gain understanding and working knowledge on a wide range of machine learning algorithms including but not limited to: linear, logistic and auto-regression models; multidimensional scaling and PCA; deep learning with multilayer perceptrons and other neural networks, support vector machines, etc.

Prerequisites	**CSCI 3070U or equivalent, MATH 2050U or equivalent**
Co-requisites	
Credit restrictions	INFR 4320U, CSCI 4160U, SOFE 3720U
Credit exemptions	
Grading scheme	__X__ **letter grade** _____ **pass/fail**

LEARNING OUTCOMES

1. Understanding of the fundamental building blocks of machine learning algorithms
2. Understanding the power and limitation of machine learning
3. Understanding the trade-offs in machine learning such as supervised vs unsupervised learning, overfitting vs regularization
4. Hands-on experience with problem solving using machine learning with software tools and programming languages
5. Working knowledge to build algorithms and systems that exhibit intelligence through machine learning

DELIVERY MODE

(check all that may apply) __X__ **face-to-face** _____ **hybrid** ____ **online**

TEACHING AND ASSESSMENT METHODS

Students are assessed by a variety of suitable methods such as:

1. Assessment with hands-on component

2. Assignments
3. Quizzes and tests and/or final examination

CONSULTATION AND FINANCIAL IMPLICATIONS, WHERE APPROPRIATE

This course is an elective with standard teaching assignment requirement (3 hr lec, 1.5 hr lab) and no significant financial implications.

APPROVAL DATES

Curriculum Committee approval	March 23, 2016
Faculty Council approval	**April 13, 2016**
Date of Submission to CPRC/GSC	**April 13, 2016**