SPECIAL REPORT

of the

ACADEMIC PRIORITIES, GRADUATE AND PROGRAM AND BUDGET COUNCILS

caring a

CERTIFICATE IN STATISTICAL AND COMPUTATIONAL DATA SCIENCE

Presented at the 758th Regular Meeting of the Faculty Senate April 28, 2016

COUNCIL MEMBERSHIP

ACADEMIC PRIORITIES COUNCIL

Faune Albert, Richard Bogartz (Chair), Nicholas Bromell, Elizabeth Chilton, Suzanne Daly, Kathleen Debevec, Jean DeMartinis, Piper Gaubatz, Bryan Harvey, Masoud Hashemi, Deborah Henson, Ayemisi Jimoh, Sangeeta Kamat, Stephen Magner, Ernest May, Katherine Newman, MJ Peterson, Monroe Rabin, James Rinderle, Peter Stern, Jack Wileden, Donna Zucker

GRADUATE COUNCIL


PROGRAM AND BUDGET COUNCIL

The College of Information and Computer Sciences and the Department of Mathematics and Statistics at the University of Massachusetts Amherst will offer the proposed Certificate in Statistical and Computational Data Science through a joint effort. The courses are currently offered through classroom learning (i.e., in person, not online). The proposed Certificate in Statistical and Computational Data Science is a five-course (15-credit) certificate that includes graduate-level course requirements from both Statistics and Computer Science. Students can choose from an approved list of courses in Statistics and Computer Science. The topics of the courses include statistical theory and methods, computational statistics, machine learning, algorithm development and data visualization. The list of approved courses is extensive so students can choose courses that align with their interests and strengths. All approved courses include skills that are needed by potential employers.

The Academic Priorities Council voted to recommend approval of the proposal at its meeting on March 24, 2016.

GRADUATE COUNCIL

Through an electronic vote conducted on March 2, 2016, the Academic Standards and Curriculum Committee (ASCC) of the Graduate Council voted to approve the Certificate in Statistical and Computational Data Science.

On Wednesday, March 9, 2016, the Graduate Council unanimously approved the Certificate in Statistical and Computational Data Science, Proposal #2017 in the Course and Curriculum Management System.

PROGRAM AND BUDGET COUNCIL

The Program Subcommittee of the Program and Budget Council met on December 9, 2015, reviewed the Certificate in Statistical and Computational Data Science proposal and recommended it for approval.

At its meeting on December 16, 2015, the Program and Budget Council unanimously approved the Certificate in Statistical and Computational Data Science, Proposal #2017 in the Course and Curriculum Management System.

MOVED: That the Faculty Senate approve the Certificate in Statistical and Computational Data Science, as presented in Sen. Doc. No. 16-056.
Briefly describe the Certificate.

The College of Information and Computer Sciences and the Department of Mathematics and Statistics at the University of Massachusetts Amherst will offer the proposed Certificate in Statistical and Computational Data Science through a joint effort. The courses are currently offered through classroom learning (i.e., in person, not online).

The proposed Certificate in Statistical and Computational Data Science is a five-course (15 credit) certificate that includes graduate-level course requirements from both Statistics and Computer Science. Students can choose from an approved list of courses in Statistics and Computer Science. The topics of the courses include statistical theory and methods, computational statistics, machine learning, algorithm development and data visualization. The list of approved courses is extensive so students can choose courses that align with their interests and strengths. All approved courses include skills that are needed by potential employers.

There is an urgent need in industry and academics for workers who are trained in data science. The purpose of the Certificate in Statistical and Computational Data Science is to prepare students to meet the need for qualified individuals trained in data science. Acquiring the certificate will help students become more competitive in the job market; the certificate also provides concrete demonstration by the student of having obtained a set of marketable skills in data science.

The required courses for the certificate are at the level of 500 and above. Graduate matriculated students and non-matriculated Continuing and Professional Education students can enroll in the required courses for the Certificate in Statistical and Computational Data Science, as long as they have the prerequisites for the courses. Courses are offered in both fall and spring. The certificate can be completed within one year.

In order to give non-computer science majors the core computer skills necessary to complete the required courses, the College of Information and Computer Sciences is developing a one-credit bridge course, COMPSCI 590NC Introduction to Numerical Computing with Python.

Provide a brief overview of the process for developing this Certificate.

The certificate was conceived because there is an almost insatiable demand for students with appropriate data science preparation and our current programs lacked the value-added emphasis on it. Since data science curriculum varies by industry and, therefore, discipline, faculty members from the Department of Biostatistics, Isenberg School of Management, College of Information and Computer Sciences, Department of Mathematics and Statistics, Department of Mechanical and Industrial Engineering and Department of Sociology met over several months to discuss how to address this need. One result of their discussions was a joint decision to pursue four separate certificates - business analytics and statistical and computational data science and two from biostatistics. For this certificate, faculty members from Computer Science and Mathematics and Statistics derived a list of courses, based on industry standards and needs expressed by potential employers.

Acknowledging that some students, e.g., non-Computer Science majors, might not have the requisite computational skills, it was agreed that a one-credit “bridge course” in “Numerical Computing in Python” would address this void. The final list of courses was identified based on course availability and capacity, and meet the requirements of a Certificate in Statistical and Computational Data Science.

Describe the Certificate’s purpose and the particular knowledge and skills that will be acquired by participating students.

The data science certificate prepares students for data-driven decision making, and deriving insights from data sets using the most up-to-date statistical methods, computational methods and analytic tools available. The program blends topics in statistical methods, statistical computing, machine learning and algorithm development to train students to become effective data scientists. The course list includes courses from Statistics and Computer Science. Additional skills that students will develop include the ability to work with large databases, and to manage and evaluate data sets and create meaningful output that can be used in effective decision making.
If this proposal requires no additional resources, say so and briefly explain why. If this proposal requires additional resources, explain how they will be paid for. For proposals involving instruction, indicate how many new enrollments are expected and whether the courses have room to accommodate them.

There is sufficient space in existing statistics and computer science courses to accommodate the expected 25 new enrollments. In addition, CICS’s tenure track and lecturer hiring plans will be providing additional instructors for the three new computer science courses (see Course Availability) and related classes.

Please describe the curriculum for this certificate, listing all required courses and possible electives, any prerequisites or GPA requirements, the recommended order or coursework and any other pertinent information. You may attach additional materials related to the curriculum at the end of this section.

Requirements for the certificate: Students must complete the required courses with a grade of "B" or better for each course. All courses must be taken through UMass Amherst. Please see the attached course listing for course descriptions and prerequisites.

Curriculum:
Required
CMPSCI 589/689 Machine Learning
One or two of the following:
CMPSCI 585/685 Introduction to Natural Language Processing
CMPSCI 590DV Data Visualization and Exploration
CMPSCI 590DS Systems for Data Science
CMPSCI 590DA Algorithms for Data Science

Two or three of the following:
STATISTIC 597A Computational Statistics
STATISTIC 697R Regression
STATISTIC 597S Introduction to Probability and Mathematical Statistics
STATISTIC 607 Mathematical Statistics I
STATISTIC 608 Mathematical Statistics II
STATISTIC 705 Linear Models

The curriculum is subject to change when appropriate by mutual consent of the Department of Mathematics and Statistics and the College of Information and Computer Sciences.

Explain how these courses represent a coherent course of study.

Data scientists need a strong foundation in statistics and computer science which can then be applied to a field of interest e.g., web services, life sciences, finance, etc. The very large quantities of data present many questions for the data scientist trying to collect, clean, or analyze it and both statistics and computer science will give him/her the tools to answer the questions or, if there are no available tools, build new tools. The statistics course choices were selected because a data scientist needs to know the underlying statistics theory as well as computational methods. The computer science courses teach the techniques to manipulate the data - algorithms, systems, natural language processing, and visualization - so that the students learn how to build usable solutions for the problems they are trying to solve. Taken together, the certificate requirements provide a rigorous course of study for aspiring data scientists.

Describe how there is a clear educational objective that can be achieved in an efficient and well-defined manner.

The educational objective of the Certificate in Statistical and Computational Data Science is for students to build statistical and computational expertise necessary to be identified by potential employers as being trained in data science techniques. The current scheduling of these courses make it possible for students to take the required five courses for the certificate within two semesters.
Explain how the course sequence offers a clear objective at the appropriate educational level.

All of the required courses are at the level of 500 and above, and are thus appropriate for the certificate at the graduate level.

Describe the perceived need for this Certificate.

Data scientists are able to bring structure to large quantities of formless data and make analysis possible. Due to the vast amounts of data that are being collected, making use of information has become key to many businesses, so much so that the demand for data scientists is outstripping the supply. The Massachusetts Technology Leadership Council has projected that there will be 120,000 data science job vacancies by 2018 in Massachusetts alone. Students are responding and have expressed avid interest in our current and proposed data science courses and programs. By explicitly grouping the classes in this certificate it allows graduate students to highlight their expertise in data science as they search for a job. And, since the certificate will be available to currently working professionals through the UMass Continuing and Professional Education Division, they can increase their knowledge about data science and job credentials without having to take a leave of absence from work to pursue a full MS degree.

If the courses that comprise the Certificate have been or currently are being offered, describe their schedule of availability. If the Certificate is comprised of new courses, describe their planned availability.

Fall
CMPSCI 585 Introduction to Natural Language Processing
CMPSCI 689 Machine Learning
STATISTIC 597A Computational Statistics
STATISTIC 607 Mathematical Statistics I
STATISTIC 697R Regression
STATISTIC 705 Linear Models

Spring
CMPSCI 589 Machine Learning
STATISTIC 597S Introduction to Probability and Mathematical Statistics
STATISTIC 608 Mathematical Statistics II

New courses:
The following Computer Science courses are being developed as part of a proposed Computer Science Masters Concentration in Data Science (proposal 2018). Some of these proposals are still in the draft phase.
Fall 2016 CMPSCI 590DS Systems for Data Science (proposal 2002)
Fall 2016 CMPSCI 590DV Data Visualization and Exploration (proposal 2004)

If the Certificate requires or includes courses from outside the sponsoring department, provide evidence of agreement(s) with the unit(s) offering those courses. You may attach any memoranda of understanding below.

See attached Memoranda of Understanding with the Department of Mathematics & Statistics. (Please refer to Proposal #2017 in the Course and Curriculum Management System.)

If the requirements for this Certificate overlap with those of another certificate or a degree or a degree program, describe that overlap.
(Note that if a student who has completed a certificate seeks clearance for a degree program that overlaps with that certificate program, the Registrar will note on the transcript that the certificate has been superseded by the degree.)

This certificate overlaps with two MS concentrations. The first is the [proposed] Masters in Computer Science with a Concentration in Computational Data Science. This certificate fulfills nearly half of its requirements. The second is the Masters in Statistics. This certificate could potentially fulfill 5 out of its 10 requirements.
What type of student is allowed to participate in this certificate program? (e.g., matriculated UMass students, non-matriculated CPE students, Five College students, graduate students, students in a specific degree program, etc.)

Students who are allowed to participate include graduate matriculated students and non-matriculated CPE students.

What role will this Certificate play in relation to other departments or degree programs on campus? Certificates vary widely across campus and may represent a subset of an existing degree program, a multidisciplinary program, or an entirely free-standing area of focus.

This certificate is a multidisciplinary program which combines the two fields of statistics and computer science.

Is this a transitional certificate program?
(Transitional certificate programs are comprised of core courses from specific degree programs and may act as stepping stones into those programs. If a student who has completed a transitional certificate matriculates to the University and completes the degree program associated with that certificate, the transcript will note that the certificate has been superseded by the degree.)

Yes
This can serve as a transitional certificate ONLY for the Masters Degree in Computer Science with a concentration in Data Science. However, the certificate is NOT transitional for the Masters Degree in Computer Science with concentrations other than Data Science; it is also NOT transitional for the Masters Degree in Statistics.

If applicable, please attach any additional material relating to the certificate (such as requirement checklists for students, etc.) below.

(Please refer to Proposal #2017 in the Course and Curriculum Management System for the Curriculum.)