

Motion to approve MS program in Data Science in CCS
Approved by the Faculty Senate
15 September 2016

Faculty Senate approved moving forward with the new MS program in Data Science in CCS.
 Faculty Senate will consider proposal again after review by NJ College and University Presidents
 and an external consultant.

PROGRAM ANNOUNCEMENT

| | |
|--|------------------------------------|
| Institution: | New Jersey Institute of Technology |
| New Program Title: | MS in Data Science |
| Degree Designation: | MS in Data Science |
| Degree Abbreviation: | MS DS |
| CIP Code and Nomenclature (<i>if possible</i>): | |
| Campus(es) where the program will be offered: | Newark Campus, NJIT |
| Date when program will begin (month and year): | January 2017 |
| List the institutions with which articulation agreements will be arranged: | None |

Is licensure required of program graduates to gain employment? Yes No

Will the institution seek accreditation for this program? Yes No

If yes, list the accrediting organization:

Descriptive Information

I. Objectives

The objective of the MS in Data Science (MSDS) program is

- to provide training in machine learning, big data analysis, and programming for data science,
- to provide training in statistical inference and probability,
- and cover applications and current topics in data science

II. Need

II.A. Need for the Program

The need comes from the fact that both private businesses and government sector are increasingly relying upon data to make decisions. These datasets are in the order of hundreds of gigabytes and require expertise not only in machine learning but also programming for data science and computing methods for big data. News articles in places like the Harvard Business Report have highlighted the need for data scientists as well as their attractive salaries. The program would satisfy this need by providing courses in machine learning, big data, and data science programming.

II.B. Relationship to the Institute Master Plans

NJIT currently offers the Big Data Management and Mining certificate that is database and data mining oriented, and a Big Data Essentials certificate that is geared towards big data. NJIT also offers an MS in Applied Statistics that lacks computer science, big data, and machine learning components, all of which are fundamental to data science.

Our proposed program is more focused on the science of data analysis that includes machine learning theory and applications, high performance and distributed computing for big datasets, programming for data science, and statistical inference. Our coursework would provide in-depth understanding of data science algorithms with applications in different areas like image recognition, bioinformatics, cyber-security, and cloud computing.

At the Institute level there is no program at NJIT that parallels the proposed MS Data Science. This new program is unique in that it covers in detail both theory and applications of the science of data analysis.

II.C. Relationship to Similar Programs in the State and Regions

New Jersey currently offers no graduate program in data science. There are certificates and programs offering data analytics and all are closely aligned with business applications.

- Rutgers University: Master of Business and Science Degree – Discovery Informatics and Data Sciences offered by Division of Continuous Education and Outreach

- Saint Peter's University: Master of Science in Data Science with a concentration in Business Analytics

In New York there are two MS Data Science programs currently being offered.

- New York University: MS Data Science offered by the Center for Data Science
- Columbia University: MS Data Science offered by the Data Science Institute.

II.D. Distinguished Programs Nationally

Nationwide (outside of New York and New Jersey) some universities offer MS degree programs aligned either with Statistics, Business, and Computer Science programs and contain a data science component.

- Stanford University: MS in Statistics: Data Science
- University of California, Berkeley: Online Master of Information and Data Science by School of Information
- Georgia Tech: Master of Science in Analytics by College of Computing
- Illinois Institute of Technology: MS Data Science by College of Science
- Indiana University: MS Data Science by School of Informatics and Computing
- Carnegie Mellon University: MS Information Systems or MS in IT with concentration Business Intelligence and Data Analytics by Heinz College
- Northwestern University: MS in Predictive Analytics by School of Continuing Studies and McCormick School of Engineering and Applied Science

III. Student Enrollments

We expect domestic students to be drawn from several different sources:

- Undergraduates at NJIT: Students completing their undergraduate studies in computing-related majors at NJIT who are interested in data science.
- Undergraduates at other institutions in the state and region: These students in computing-related majors, who are interested in data science, will be provided a public university program in data science in the state and region.
- Working professionals: There is a large regional pool of professionals in the NY-NJ area with jobs related to data science who are potentially interested to pursue an MS data science.

We also expect to recruit foreign students, who will find the combination of high quality education in data science and access to a large job market attractive.

We anticipate an enrollment of approximately 30 new students in the first year, increasing to 60-80 students over the next four years. This is based upon the existing enrollments in core courses currently being taught.

The College of Computing Sciences will continue to work with both Graduate Admissions and the Murray Center for Women in Technology to target women and minority students.

Admission Requirements. To be eligible for admission, a student must have an undergraduate degree with a minimum GPA of 3.0 on a 4.0 scale and have completed the following undergraduate coursework:

- Calculus I and II (equivalent to Math 111 and Math 112)
 - Derivatives, integrals, applications
 - Business calculus may suffice and will be considered on a case by case basis
- Introduction to programming (equivalent to CS 113)
 - Basic programming constructs, writing and debugging programs, iteration, recursion, arrays, lists
- Data structures and algorithms (equivalent to CS 114)
 - Basic data structures (lists, arrays, hash tables), search and sort, algorithm analysis
- Probability and statistics (equivalent to Math 333)
 - Random variables, probability distributions, sample mean and variance
 - Basic probability or statistics course separately will also suffice
- Linear algebra (equivalent to Math 337)
 - Vector spaces, dot products, Euclidean norm, matrices

International students will have to take TOEFL and GRE exams and meet minimum requirements for admission to graduate programs at NJIT.

Students who don't meet these requirements but already have an existing degree in a technical scientific subject will be evaluated on a case-by-case basis.

IV. Resources to Support the Program

Core courses are taught by Computer Science and Math faculty whereas electives are covered by faculty in their respective departments. Laboratory and equipment requirements are described under IV.D.

IV.A. Course Development

The following new course will need to be developed to complete the curriculum:

CS 6xx: Current topics in data science

Description: This course will cover current topics in data science such as deep learning, gradient boosting, GPU programming, representation learning, stacking, and current applications. Each student will present two recent papers in data science and undertake a hands-on project.

IV.B. Faculty

Faculty members from the Department of Computer Science will support the majority of the educational effort for this program. Together, they will provide extensive expertise in various aspects of data science.

IV.C. Libraries and Computing Facilities

Since this program will draw upon existing courses and the same supplemental literature that supports other related NJIT programs, library holdings are more than adequate to support the new program. NJIT's Van Houten Library has a collection of more than 130,000 books and subscribes to about 500 printed journals and about 13,000 electronic journals. The library's home page provides access to the library's online catalog and links to a wide array of information services. The library purchases between 2,500 and 3,000 new books each year. Requests for new books or journals are made through the academic department's faculty representative to the library.

The library has a wide array of networked PCs that provide access to a large number of bibliographical databases and full-text electronic journals, for searching Internet sites, searching each library's on-line catalog, and access to various on-line journal databases. VCR's for viewing videocassettes reserved for courses are also available. Journal and conference literature in engineering, science, management, architecture and other subject areas is accessible through a variety of indexing and abstracting databases. Among the databases available on line are CompendexWeb (Engineering Index); ProQuest Direct (articles on business, management and industry), Applied Science and Technology Index. The library also borrows through interlibrary loan (ILL) for materials.

The libraries web site describes the services and resources more completely. Please see <http://www.library.njit.edu>.

As a technological research university, NJIT has excellent computing systems, networks and software to support this program. The Newark campus' gigabit Ethernet network backbone connects more than 6,000 nodes in classrooms, laboratories, residence halls, faculty and staff offices, the library, and student organization offices. Wireless access is available in over 90% of campus buildings and locations. The network provides access to a wealth of shared information services. Some of these include high-performance computing servers providing CPU cycles for simulation and computational research, disk arrays for storage of large data sets, communication servers for electronic mail and document exchange, databases, digital journal subscriptions and a virtual "Help Desk." A virtual private network combined with Internet access, plus a large ISDN modem bank extend access to campus information resources to faculty, staff and students working at home, work, any of the university's extension sites or throughout the world. Wide-area network access through NJEdge.Net (New Jersey's Higher Education Network) and the Internet provide collaboration opportunities with students, faculty, and researchers, locally, regionally, nationally, and throughout the world.

IV.D. Classrooms and Laboratories

There is a broad range of classrooms and laboratories available to offer the courses in the proposed program, including many with Internet access and multimedia facilities. The university's high performance infrastructure provides GPU computing nodes and a large cluster of general purpose computing node that would facilitate big data training.

V. Curriculum (on a separate page)

The Department of Computer Science will administer the MS in Data Science. A CS faculty committee composed of CS faculty and lecturers together with CS administrative staff will oversee advisement of students in the program and monitor their progress. The Faculty Advisory Committee for the program will consist of the following CS faculty:

U. Roshan
Z. Wei
Chase Wu
Loh Ji Meng (from Math)

The following Administrative Staff will assist the Faculty Advisory Committee:

George Olsen (MS Advisor)

The courses supporting the MSCSP program are divided in three categories: (i) Core (ii) Elective, and (iii) Foundational. An MSCSP student may require the background represented by Foundational courses in order to prepare for the Core and Elective courses.

An MSCSP course program must satisfy the following distribution requirement:

- 30 credits are required, which can be satisfied as either one of the following options:
 - o Courses (30 credits)
 - o Courses (27 credits) + MS Project (3 credits)
 - o Courses (24 credits) + MS Thesis (6 credits)
- All core courses are required.
- At most three courses can be chosen from outside the Department of Computer Science with the approval of advisors

If a student chooses the MS project or MS thesis option, the project or thesis must be related to data science.

Students will be given a choice of a computational or applied statistics track. Both tracks share the same admissions requirements and electives but have a different set of core courses as given below.

Core Courses

Computational track:

- CS 675 – Machine learning
 - New course taught in Fall 2013 as special topic with enrollment of 14
 - Taught in Fall 2014 with enrollment of 23
 - Taught in Fall 2015 with enrollment of 35
 - Pre-reqs: Prior programming and basic undergraduate probability and linear algebra. Otherwise approval of instructor is required.
- CS 644 – Big data
 - New course taught in Fall 2015 with enrollment of 40
 - Taught in Spring 2016 with enrollment of 37
 - Pre-reqs: CS 435 or approval of instructor.
- CS 636 – Data analytics with R programming
 - New course taught as special topic in Spring 2016 with enrollment of 30
 - Pre-reqs: CS 280 and Math 333 or approval of instructor.
- CS 6xx – Current topics in data science
 - New course to be taught in Spring 2017
 - Pre-reqs: CS 675 or approval of instructor
- Math 661 – Applied Statistics
 - Pre-req is Math 112 (Calculus II)

Statistics track:

- Math 660 – Introduction to statistical computing with SAS and R
 - Pre-reqs: Basic knowledge in statistical concepts or approval of instructor
- Math 662 – Probability distributions
 - Pre-reqs: MATH 341 or MATH 333
- Math 665 – Statistical inference
 - Pre-reqs: MATH 662
- CS 644 – Big data
 - New course taught in Fall 2015 with enrollment of 40
 - Taught in Spring 2016 with enrollment of 37
 - Pre-reqs: CS 435 or approval of instructor.
- CS 675 – Machine learning
 - New course taught in Fall 2013 as special topic with enrollment of 14
 - Taught in Fall 2014 with enrollment of 23
 - Taught in Fall 2015 with enrollment of 35
 - Pre-reqs: Prior programming and basic undergraduate probability and linear algebra. Otherwise approval of instructor is required.

Electives

Students would have to take required pre-reqs or seek approval of instructor for the elective courses. Some electives are also core courses in one of the above tracks but not the other.

- CS 610 Data structures and algorithms
- CS 631 Data management system design

- CS 632 Advance database system design
- CS 634 Data mining
- CS 636 Data analytics with R programming
- CS 639 Electronic medical records
- CS 643 Cloud computing
- CS 645 Security and privacy in computer systems
- CS 659 Image processing and analysis
- CS 670 Artificial intelligence
- CS 683 Software project management
- CS 684 Software testing and reliability
- CS 681 Computer vision
- CS 661 Systems simulation
- CS 708 Advance data security and privacy
- CS 731 Application of database systems
- CS 732 Advance machine learning
- CS 735 High performance analytics for data science
- CS 744 Data mining and management in bioinformatics
- CS 782 Pattern recognition
- Math 630 Linear algebra and applications
- Math 631 Linear algebra
- Math 644 Regression analysis methods
- Math 660 Introduction to statistical computing with SAS and R
- Math 664 Methods for statistical consulting
- Math 665 Statistical inference
- Math 678 Introduction to statistical methods in data science
- Math 699 Design and analysis of experiments
- Math 717 Inverse problems and global optimization
- Math 786 Large sample theory and inference
- Math 787 Non-parametric statistics
- BIOL 638 Computational ecology
- MGMT 635 Data mining and analytics for managers
- MGMT 630 Decision analysis
- FIN 600 Corporate finance I
- FIN 641 Derivatives markets
- FIN 642 Derivatives and structured finance
- MRKT 630 Models of consumer behavior
- MRKT 613 Market planning and analysis
- IS 665 Data analytics for information systems
- IS 631 Enterprise database management
- IS 688 Web mining
- IS 687 Transaction mining and fraud detection
- BNFO 601 Foundations of bioinformatics I
- BNFO 602 Foundations of bioinformatics II
- BNFO 615 Data analysis for bioinformatics

- BNFO 620 Genomic data analysis

Keywords: MS, Data Science, CCS, College of Computing Sciences