University Senate

Proposed: November 15, 2013
Adopted: November 15, 2013 without dissent

RESOLUTION TO APPROVE A MASTER OF SCIENCE IN DATA SCIENCE (SEAS)

WHEREAS, as data collection systems continue to improve, large aggregations of data can increasingly be mined to tackle the world's most challenging problems, and

WHEREAS, data science has applications in medicine, infrastructure monitoring, finance, social media, data security, and myriad other fields, and

WHEREAS, there is a shortage of skilled analysts and managers capable of harnessing the power of big data, and

WHEREAS, major corporations have approached the university about graduate training opportunities in data science, and

WHEREAS, there is considerable student demand for such a program, and

WHEREAS, the proposed program includes a core encompassing machine learning, statistical modeling, computer systems and data visualization, and

WHEREAS, the proposed program offers students electives across the university tailored to their career objectives, and

WHEREAS, the program's offerings are drawn from the Graduate School of Arts and Sciences' Department of Statistics, the Departments of Computer Science and Industrial Engineering and Operations Research in the School of Engineering and Applied Science, and the Institute for Data Sciences and Engineering, and

WHEREAS, the proposed program does not replace or duplicate any existing program at the university,

NOW, THEREFORE, BE IT RESOLVED that the Senate approve the program leading to a Master of Science in Data Science.

BE IT FURTHER RESOLVED that the resolution be forwarded to the Trustees of Columbia University for their approval.

Proponent
Committee on Education

Data is everywhere in today's world. Everything can be sensed, from the human body to the environment, adding to the exabytes of online and mobile data consumed each month. The Institute for Data Sciences and Engineering will develop technology to unlock the power of global data to solve some of society's most challenging problems. The purpose of the Master of Science in Data Science program curriculum is to assist in developing the tools and expertise to (as examples): monitor critical infrastructure, improve patient care, extract information from digital footprints, enable social interaction, track market trades, and keep personal data safe. In response to New York City's Applied Science initiative, and drawing upon the academic strengths across Columbia University, this program will produce talented graduates armed with the theories and knowledge they need to apply data science to society's most challenging problems. The Master of Science program includes a comprehensive core aimed at building expertise in machine learning, statistical modeling, computer systems and data visualization, coupled with a set of application-oriented electives that fall into tracks corresponding to the six Centers of the Institute for Data Sciences and Engineering, as well as an Entrepreneurship concentration. Students will use the technical skills acquired through their coursework to develop tools and solutions through a required capstone and ethics course.

The Education Committee of the Institute for Data Sciences and Engineering (IDSE) designed the 30-credit master's-level program. The program's core of six courses (18 credits) will be offered by the Department of Statistics within the Graduate School of Arts and Sciences (GSAS) and the Computer Science and Industrial Engineering and Operations Research Departments within The Fu Foundation School of Engineering and Applied Science (SEAS). The program's capstone and ethics course (3 credits) will be a collaboration between faculty who might be drawn from all departments in SEAS and the IDSE's industrial affiliates program. The three courses (9 credits) of technical elective coursework will draw upon the strength of the graduate-level course offerings across the University to tailor to the students' particular career interests, in collaboration with faculty advisement. The institute's Executive Committee, SEAS and GSAS have granted approval for the program, including its structure and purpose and various aspects of its administration.

The proposed M.S. in Data Science includes the following existing courses: Algorithms for Data Science (CSOR), Machine Learning for Data Science (COMS), and Exploratory Data Analysis and Visualization (STATS), as well as three new courses offered by SEAS and GSAS: Computer Systems for Data Science (COMS), Statistical Inference and Modeling (STATS), and Data Science Capstone and Ethics. The seventh core course strengthens the knowledge in probability and statistics that the program delivers by drawing upon an existing course within the Master of Arts in Statistics, Probability.

In keeping with the institute's mission, the M.S. in Data Science program is designed to be interdisciplinary with the graduate elective coursework (3 courses, 9 credits) spanning the University to create a diverse technical curriculum.

There are six Centers within the IDSE. Each will provide a source of technology projects associated with important global issues. These projects will be used for mini-projects or exercises within the core course or as the basis for a capstone project that also engages industry. Beyond the core curriculum, as students select their graduate elective course, they will have an option to choose from a set of application-oriented concentrations whose

courses align with the intellectual thrusts of the six centers: Cybersecurity, Financial Analytics, Foundation of Data Science, Health Analytics, New Media and Smart Cities. Additionally, to embody the Institute's focus on technology and innovation, students will have the option to pursue a concentration of electives in Entrepreneurship. Finally, to provide flexibility given the seven-course required core, students can also fulfill their three technical electives by selecting courses that cut across the Center and Entrepreneurship concentrations, or propose their own electives. The electives will draw upon existing graduate-level courses at Columbia. Students will require advisor approval when selecting electives, and the choice will be subject to course prerequisites, course availability and cross-registration procedures of the school or department offering the requested courses.

In general, completed coursework will not be eligible to count for more than one program as outlined by Columbia University policies. Master of Science students who enter the program having completed prior related coursework at Columbia University will be bound by the advanced standing policies as outlined by the SEAS bulletin.