



Data Science

The Master of Science in Data Science is a 30-credit program that allows students to apply data science techniques to their field of interest. Our students have the opportunity to conduct original research, included in a capstone project, and interact with our industry partners and faculty. Students are required to complete 21-credits of core coursework and 9-credits of electives.

Core Courses

CSOR W4246 ALGORITHMS FOR DATA SCIENCE (3-credits)

Methods for organizing data, e.g. trees, queues, lists, priority queues. Sorting and searching. Basic graph models and algorithms for searching and shortest paths. Hashing and randomized algorithms and data structures for answering questions on big data. Data compression techniques. Greedy algorithms and dynamic programming. Network flows and assignment problems. Linear and integer programming. NP-completeness. Streaming algorithms for computing statistics on the data.

COMS W4121 COMPUTER SYSTEMS FOR DATA SCIENCE (3-credits)

An introduction to computer architecture and distributed systems with an emphasis on warehouse scale computing systems. Topics will include fundamental tradeoffs in computer systems, hardware and software techniques for exploiting instruction-level parallelism, data-level parallelism and task level parallelism, scheduling, caching, prefetching, network and memory architecture, latency and throughput optimizations, specialization, and an introduction to programming data center computers.

COMS W4721 MACHINE LEARNING FOR DATA SCIENCE (3-credits)

A graduate-level introduction to machine learning. The course covers basic statistical principles of supervised machine learning, as well as some common algorithmic paradigms. Additional topics, such as representation learning and online learning, may be covered if time permits.

STAT GR5701 PROBABILITY AND STATISTICS FOR DATA SCIENCE (3-credits)

This course covers the following topics: Fundamentals of probability theory and statistical inference used in data science; Probabilistic models, random variables, useful distributions, expectations, law of large numbers, central limit theorem; Statistical inference; point and confidence interval estimation, hypothesis tests, linear regression.

STAT GR5702 EXPLORATORY DATA ANALYSIS AND VISUALIZATION (3-credits)

Fundamentals of data visualization, layered grammar of graphics, perception of discrete and continuous variables, introduction to Mondrian, mosaic plots, parallel coordinate plots, introduction to ggobi, linked plots, brushing, dynamic graphics, model visualization, clustering and classification.

STAT GR5703 STATISTICAL INFERENCE AND MODELING (3-credits)

Course covers fundamentals of statistical inference and testing, and gives an introduction to statistical modeling. The first half of the course will be focused on inference and testing, covering topics such as maximum likelihood estimates, hypothesis testing, likelihood ratio test, Bayesian inference, etc. The second half of the course will provide introduction to statistical modeling via introductory lectures on linear regression models, generalized linear regression models, nonparametric regression, and statistical computing. Throughout the course, real-data examples will be used in lecture discussion and homework problems.

ENGI E4800 DATA SCIENCE CAPSTONE AND ETHICS (3-credits)

This course provides a unique opportunity for students in the M.S. in Data Science program to apply their knowledge of the foundations, theory and methods of data science to address data science problems in industry, government and the non-profit sector. The course activities focus on a semester-length data science project sponsored by a faculty member or local organization. The project synthesizes the statistical, computational, engineering challenges and social issues involved in solving complex real-world problems.



COLUMBIA | ENGINEERING

The Fu Foundation School of Engineering and Applied Science

Data Science Faculty



Eleni Drinea - MS Program Director

The Fu Foundation School of Engineering and Applied Science

Senior Lecturer in the Discipline of Computer Science and in

the Discipline of Industrial Engineering and Operations

Research



Marco Avella

Faculty of Arts and Sciences

Assistant Professor



Asaf Cidon

The Fu Foundation School of Engineering and Applied Science

Assistant Professor of Electrical Engineering



John Paisley

The Fu Foundation School of Engineering and Applied Science

Associate Professor of Electrical Engineering



Joyce Robbins

Faculty of Arts and Sciences

Lecturer in the Discipline of Statistics



Sining Chen

The Fu Foundation School of Engineering and Applied Science

Adjunct Professor of Industrial Engineering and Operations

Research



Tat Sang Fung

Faculty of Arts and Sciences

Adjunct Assistant Professor



Adam Kelleher

The Fu Foundation School of Engineering and Applied Science

Adjunct Professor of Computer Science