Data Science

OUR MISSION

To prepare students for a career in Data Science; to educate them in the scientific foundations and methodologies to understand, explore, process, and interpret data; to train them to apply data scientific tools to meet the challenges of the future; to promote a sense of scholarship, leadership and service among our graduates; to instill in the students the desire to create, develop, and disseminate new knowledge; and to produce international leaders in data science and its related professions.

MASTER OF SCIENCE IN DATA SCIENCE

The Master of Science in Data Science program provides students from a variety of backgrounds with a strong technical education in data scientific concepts and tools so that they may create innovative solutions to address societal challenges using data, state-of-the-art analytical methods and computing technology. Graduates from the program will gain proficiency required for positions in research and development within data science and its application in a variety of fields, and have the academic training to pursue doctoral research in or using data science.

Full-time students can complete the 30-credit program in as little as 11 months; part-time students may require up to 3 years. DSCI courses are available in person or online. A program of study must be submitted in compliance with college regulations.

Program Requirements

The program consists of:

DSCI Core Courses	21
Approved Electives	9
Total Credits	30

The seven required DSCI core courses are:

DSCI 310	Introduction to Data Science	3
DSCI 311	Optimization and Mathematical Foundations for Data Science	3
DSCI 321	Algorithms and Software Foundations for Data Science	3
DSCI 411 or DSCI 421	Data Management for Big Data Accelerated Computing for Machine Learning	3
DSCI 431	Introduction to Statistical Modeling	3
DSCI 441	Statistical and Machine Learning	3
DSCI 451	Ethics in Data Science	3

In addition to the core requirements, students are required to complete a minimum of 9 credits from a list of approved electives on the program website, at least 6 of which must be at the 400 level, and can optionally include up to six credits of thesis work. At most 3 courses (totaling 9 credits) from other programs can be applied towards the requirements of this program.

Recommended sequence of courses (1-Year ACCELERATED PROGRAM)

Summer Session II (July/August)

DSCI 310	Introduction to Data Science	3
DSCI 311	Optimization and Mathematical Foundations for Data Science	3

Fall Semester

DSCI 321	Algorithms and Software Foundations for Data Science	3
DSCI 431	Introduction to Statistical Modeling	3
DSCI 451	Ethics in Data Science	3
Approved Elective		3

Spring Semester

DSCI 411 or DSCI 421	Data Management for Big Data Accelerated Computing for Machine Learning	3
DSCI 441	Statistical and Machine Learning	3
Approved Elective		3
Approved Elective		3

RECOMMENDED SEQUENCE OF COURSES (1.5-YEAR PROGRAM)

Year 1 Summer Session II (July/August)

DSCI 311	Optimization and Mathematical Foundations for Data Science	3

Year 1 Fall Semester

DSCI 310	Introduction to Data Science	3
DSCI 321	Algorithms and Software Foundations for Data Science	3
DSCI 431	Introduction to Statistical Modeling	3

Year 1 Spring Semester

DSCI 411 or DSCI 421	Data Management for Big Data Accelerated Computing for Machine Learning	3
DSCI 441	Statistical and Machine Learning	3
Approved Elective		3

Year 2 Fall Semester

DSCI 451	Ethics in Data Science	3
Approved Elective		3
Approved Elective		3

GRADUATE CERTIFICATE IN DATA SCIENCE

The Graduate Certificate in Data Science program provides students with an introduction to the basic concepts and tools in data science. Individuals completing this program will be better positioned to understand and explore the application of data scientific concepts and methodologies in a variety of domains, or pursue more advanced training in Data Science or a field that requires a data scientific skillset. Upon completion of the certificate program, students can further enhance their knowledge and skills in the field by applying to the Master of Science in Data Science degree program and applying the 12 certificate credits towards the 30-credit Master's degree.

REQUIRED COURSES

Four courses (12 credits) are required in total for the certificate. Select two of the following 300-level fundamentals of data science courses (3 credits each):

DSCI 310	Introduction to Data Science	3
DSCI 311	Optimization and Mathematical Foundations for Data Science	3
DSCI 321	Algorithms and Software Foundations for Data Science	3

The remaining two courses are at the 400-level. Select one or two of the following 400-level core data science courses (3 credits each), and at most one course from a list of approved alternatives (also at the 400 level). Note that DSCI 431 and ECE 414 cannot both be chosen.

DSCI 411	Data Management for Big Data	3
DSCI 421	Accelerated Computing for Machine Learning	3
DSCI 431	Introduction to Statistical Modeling	3
DSCI 441	Statistical and Machine Learning	3
DSCI 451	Ethics in Data Science	3

Approved alternative courses include:

CSE 425	Natural Language Processing	3
CSE 447	Data Mining	3

CSE 449	Big Data Analytics	3
ECE 414	Statistical Decision Making and Machine Learning Theory	3
ECE 440	Introduction to Online and Reinforcement Learning	3
ISE 409	Time Series Analysis	3
ISE 410	Design of Experiments	3
ISE 444	Optimization Methods in Machine Learning	3
ISE 465	Applied Data Mining	3
STAT 439	Time Series and Forecasting	3

Generally, the 400-level courses will have prerequisites such that the 300-level courses are taken first, but there is no prescribed order for the courses.

An undergraduate minor in data science is offered within the Computer Science and Engineering Department.

Courses

DSCI 301 Mathematics for Data Science 3 Credits
Concepts from multivariable calculus, linear algebra/methods, statistics and probability as useful in a data science context. Course may not be taken for credit toward the MS in Data Science but can satisfy prerequisites.
Prerequisites: MATH 022 or MATH 032

DSCI 310 Introduction to Data Science 3 Credits
The computational analysis of data to extract knowledge and insight. Exploration and manipulation of data. Introduction to data collection and cleaning, reproducibility, code and data management, statistical inference, modeling, ethics, and visualization. Not available to undergraduate students.
Prerequisites: CSE 004 or CSE 007 or CSE 012 or BIS 335

DSCI 311 Optimization and Mathematical Foundations for Data Science 3 Credits
Introduction to optimization for data science. Topics in mathematical structures, linear modeling and matrix computation, and probabilistic thinking and modeling.
Prerequisites: DSCI 301

DSCI 321 Algorithms and Software Foundations for Data Science 3 Credits
Foundational computer science topics and software development in Python for data science. Concepts from discrete structures, algorithm design, programming concepts and data structures, object-oriented programming, exception handling, tools and environments, and scaling for big data.
Prerequisites: (CSE 004 or CSE 007 or CSE 012 or BIS 335) and (MATH 021 or MATH 031 or MATH 076)

DSCI 392 Independent Study 1-3 Credits
An intensive study, with report, of a topic in data science which is not treated in other courses. Consent of instructor required.
Repeat Status: Course may be repeated.

DSCI 411 Data Management for Big Data 3 Credits
Modern distributed systems for big data. Systems and technology such as SQL, NoSQL, Hadoop, Spark. Data collection, cleaning, structuring and transforming data, data provenance.
Prerequisites: DSCI 310 and DSCI 321

DSCI 421 Accelerated Computing for Machine Learning 3 Credits
Introduction to hardware architectures and parallel computing systems that facilitate high speed machine learning. Graphics processing units (GPUs), hardware architecture of parallel computers, memory allocation and data parallelism, multidimensional kernel configuration, kernel-based parallel programming, principles and patterns of parallel algorithms, application of parallel computing to machine learning.
Prerequisites: DSCI 310 and DSCI 321

DSCI 431 Introduction to Statistical Modeling 3 Credits
Statistical analysis of data and linear models. Exploratory data analysis, graphical data analysis, estimation and hypothesis testing, Bayesian methods, simulation and resampling, linear, multivariate and generalized linear models, model selection and performance evaluation.
Prerequisites: DSCI 310 and DSCI 311

DSCI 441 Statistical and Machine Learning 3 Credits
Common machine learning methods, algorithmic analysis of models for scalability and implementation, data transformations (including dimension reduction, smoothing, aggregation), supervised and unsupervised learning, and ensemble methods.
Prerequisites: DSCI 310 and DSCI 321 and DSCI 431

DSCI 451 Ethics in Data Science 3 Credits
Legal and ethical considerations including privacy, reproducibility, bias, and fairness that are central to data science efforts, as well as ethical principles in information and technology research. Issues in real-world contexts. Development of technical solutions.
Prerequisites: DSCI 310 and DSCI 321

DSCI 480 Capstone Experience 3 Credits
Design, implementation, and evaluation of a data science project. Small student teams. Project definition, planning, data acquisition, analysis, evaluation, and documentation. Communication skills such as technical writing, oral presentation, and visualization.
Prerequisites: DSCI 311 and (DSCI 411 or DSCI 421)
Corequisites: DSCI 441 and DSCI 451

DSCI 490 Thesis 1-6 Credits
Thesis. Permission required.
Repeat Status: Course may be repeated.

DSCI 492 Independent Study 1-3 Credits
An intensive study, with report, of a topic in data science which is not treated in other courses. Consent of instructor required.
Repeat Status: Course may be repeated.