

## Data Science (DSCI)

---

### 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.