

Data Science (DATASCI)

Please note: DATASCI courses are only available for Information and Data Science (MIDS) students.

Expand all course descriptions [+] Collapse all course descriptions [-]

DATASCI W200 Introduction to Data Science Programming 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

This fast-paced course gives students fundamental Python knowledge necessary for advanced work in data science. Students gain frequent practice writing code, building to advanced skills focused on data science applications. We introduce a range of Python objects and control structures, then build on these with classes on object-oriented programming. A major programming project reinforces these concepts, giving students insight into how a large piece of software is built and experience managing a full-cycle development project. The last section covers two popular Python packages for data analysis, Numpy and Pandas, and includes an exploratory data analysis.

Introduction to Data Science Programming: Read More [+]

Objectives & Outcomes

Student Learning Outcomes: Be able to design, reason about, and implement algorithms for solving computational problems.
Be able to generate an exploratory analysis of a data set using Python.
Be able to navigate a file system, manipulate files, and execute programs using a command line interface.
Be able to test and effectively debug programs.
Be fluent in Python syntax and familiar with foundational Python object types.
Be prepared for further programming challenges in more advanced data science courses.
Know how to read, manipulate, describe, and visualize data using the Numpy and Pandas packages.
Know how to use Python to extract data from different type of files and other sources.
Understand how to manage different versions of a project using Git and how to collaborate with others using Github.
Understand the principles of functional programming.
Understand the principles of object-oriented design and the process by which large pieces of software are developed.

Rules & Requirements

Prerequisites: MIDS students only

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Laskowski

Introduction to Data Science Programming: Read Less [-]

DATASCI W201 Research Design and Applications for Data and Analysis 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

Introduces the data sciences landscape, with a particular focus on learning data science techniques to uncover and answer the questions students will encounter in industry. Lectures, readings, discussions, and assignments will teach how to apply disciplined, creative methods to ask better questions, gather data, interpret results, and convey findings to various audiences. The emphasis throughout is on making practical contributions to real decisions that organizations will and should make. Course must be taken for a letter grade to fulfill degree requirements.

Research Design and Applications for Data and Analysis: Read More [\[+\]](#)

Rules & Requirements

Prerequisites: Master of Information and Data Science students only

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Weber

Research Design and Applications for Data and Analysis: Read Less [\[-\]](#)

DATASCI W203 Statistics for Data Science 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

This course provides students with a foundational understanding of classical statistics within the broader context of data science. Topics include exploratory analysis and descriptive statistics, probability theory and the foundations of statistical modeling, estimators, hypothesis testing, and classical linear regression. Causal inference and reproducibility issues are treated briefly. Students will learn to apply the most common statistical procedures correctly, checking assumptions and responding appropriately when they appear violated; to evaluate the design of a study and how the variables being measured relate to research questions; and to analyze real-world data using the open-source language R.

Statistics for Data Science: Read More [\[+\]](#)

Rules & Requirements

Prerequisites: Master of Information and Data Science students only.

Intermediate competency in calculus is required. A college-level linear algebra course is recommended

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Cheshire

Statistics for Data Science: Read Less [\[-\]](#)

DATASCI W205 Fundamentals of Data Engineering 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

Storing, managing, and processing datasets are foundational processes in data science. This course introduces the fundamental knowledge and skills of data engineering that are required to be effective as a data scientist. This course focuses on the basics of data pipelines, data pipeline flows and associated business use cases, and how organizations derive value from data and data engineering. As these fundamentals of data engineering are introduced, learners will interact with data and data processes at various stages in the pipeline, understand key data engineering tools and platforms, and use and connect critical technologies through which one can construct storage and processing architectures that underpin data science applications.

Fundamentals of Data Engineering: Read More [+]

Rules & Requirements

Prerequisites: Master of Information and Data Science students only. Intermediate competency in Python, C, or Java, and competency in Linux, GitHub, and relevant Python libraries; or permission of instructor. Knowledge of database management including SQL is recommended but not required

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of web-based lecture per week

Summer: 15 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructors: Mims, Martin

Fundamentals of Data Engineering: Read Less [-]

DATASCI W207 Applied Machine Learning 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

Machine learning is a rapidly growing field at the intersection of computer science and statistics concerned with finding patterns in data. It is responsible for tremendous advances in technology, from personalized product recommendations to speech recognition in cell phones. This course provides a broad introduction to the key ideas in machine learning. The emphasis will be on intuition and practical examples rather than theoretical results, though some experience with probability, statistics, and linear algebra will be important. Course must be taken for a letter grade to fulfill degree requirements.

Applied Machine Learning: Read More [+]

Rules & Requirements

Prerequisites: Master of Information and Data Science students only. Data Science W201, W203. Intermediate competency in Python, C, or Java, and competency in Linux, GitHub, and relevant Python libraries; or permission of instructor. Linear algebra is recommended

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Hearst

Applied Machine Learning: Read Less [-]

DATASCI W209 Data Visualization 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

Visualization enhances exploratory analysis as well as efficient communication of data results. This course focuses on the design of visual representations of data in order to discover patterns, answer questions, convey findings, drive decisions, and provide persuasive evidence. The goal is to give you the practical knowledge you need to create effective tools for both exploring and explaining your data. Exercises throughout the course provide a hands-on experience using relevant programming libraries and software tools to apply research and design concepts learned.

Data Visualization: [Read More](#) [+]

Objectives & Outcomes

Student Learning Outcomes: Analyze data using exploratory visualization.

Build commonly requested types of visualizations as well as more advanced visualizations using ground-up customization.

Constructively critique existing visualizations, identifying issues of integrity as well as excellence.

Create useful, performant visualizations from real-world data sources, including large and complex datasets.

Design aesthetically pleasing static and interactive visualizations with perceptually appropriate forms and encodings.

Improve your own work through usability testing and iteration, with attention to context.

Select appropriate tools for building visualizations, and gain skills to evaluate new tools.

Rules & Requirements

Prerequisites: Master of Information and Data Science students only. DATASCI W203. Students must take DATASCI W205 concurrently or prior to DATASCI W209. If taken concurrently, students may not drop W205 and remain in W209. Recommended: experience with HTML, CSS, and JavaScript, or ability to learn new programming languages quickly. If Python is the only programming language you know, you will probably benefit from learning the basics of web development with JavaScript in advance

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Data Visualization: [Read Less](#) [-]

DATASCI W210 Capstone 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

The capstone course will cement skills learned throughout the MIDS program – both core data science skills and “soft skills” like problem-solving, communication, influencing, and management – preparing students for success in the field. The centerpiece is a semester-long group project in which teams of students propose and select project ideas, conduct and communicate their work, receive and provide feedback (in informal group discussions as well as formal class presentations), and deliver compelling presentations along with a Web-based final deliverable. Includes relevant readings, case discussions, and real-world examples and perspectives from panel discussions with leading data science experts and industry practitioners.

Capstone: [Read More](#) [+]

Rules & Requirements

Prerequisites: Students must be in their final semester of the MIDS program

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Capstone: [Read Less](#) [-]

DATASCI W231 Behind the Data: Humans and Values 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

Intro to the legal, policy, and ethical implications of data, including privacy, surveillance, security, classification, discrimination, decisional-autonomy, and duties to warn or act. Examines legal, policy, and ethical issues throughout the full data-science life cycle — collection, storage, processing, analysis, and use — with case studies from criminal justice, national security, health, marketing, politics, education, employment, athletics, and development. Includes legal and policy constraints and considerations for specific domains and data-types, collection methods, and institutions; technical, legal, and market approaches to mitigating and managing concerns; and the strengths and benefits of competing and complementary approaches.

Behind the Data: Humans and Values: [Read More](#) [+]

Rules & Requirements

Prerequisites: MIDS and MPA students only

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Mulligan

Behind the Data: Humans and Values: [Read Less](#) [-]

DATASCI W233 Privacy Engineering 3 Units

Terms offered: Not yet offered

This course surveys privacy mechanisms applicable to systems engineering, with a particular focus on the inference threat arising due to advancements in artificial intelligence and machine learning. We will briefly discuss the history of privacy and compare two major examples of general legal frameworks for privacy from the United States and the European Union. We then survey three design frameworks of privacy that may be used to guide the design of privacy-aware information systems. Finally, we survey threat-specific technical privacy frameworks and discuss their applicability in different settings, including statistical privacy with randomized responses, anonymization techniques, semantic privacy models, and technical privacy mechanisms.

Privacy Engineering: [Read More](#) [+]

Rules & Requirements

Prerequisites: MIDS students only

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Privacy Engineering: [Read Less](#) [-]

DATASCI W241 Experiments and Causal Inference 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

This course introduces students to experimentation in the social sciences.

This topic has

increased considerably in importance since 1995, as researchers have learned to think

creatively about how to generate data in more scientific ways, and developments in information

technology have facilitated the development of better data gathering. Key to this area of inquiry is

the insight that correlation does not necessarily imply causality. In this course, we learn how to

use experiments to establish causal effects and how to be appropriately skeptical of findings

from observational data.

Experiments and Causal Inference: Read More [+]

Rules & Requirements

Prerequisites: Data Science W201 and W203

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of web-based lecture per week

Summer: 15 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Experiments and Causal Inference: Read Less [-]

DATASCI W251 Deep Learning in the Cloud and at the Edge 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

This hands-on course introduces data scientists to technologies related to building and operating live, high throughput Deep Learning applications running on powerful servers in the Cloud as well on smaller and lower power devices at the Edge of the Network. The material of the class is a set of practical approaches, code recipes, and lessons learned. It is based on the latest developments in the Industry and industry use cases as opposed to pure theory. It is taught by professionals with decades of industry experience.

Deep Learning in the Cloud and at the Edge: Read More [+]

Rules & Requirements

Prerequisites: Master of Information and Data Science students only. Students must have completed Data Science W201, W203, and W205 before enrolling in this course. They should be able to program in C, Python, or Java and/or be able to pick up a new programming language quickly. A degree of fluency is expected with the basics of operating systems (e.g., Linux and the Internet Technologies

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Deep Learning in the Cloud and at the Edge: Read Less [-]

DATASCI W261 Machine Learning at Scale 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

This course teaches the underlying principles required to develop scalable machine learning pipelines for structured and unstructured data at the petabyte scale. Students will gain hands-on experience in Apache Hadoop and Apache Spark.

Machine Learning at Scale: Read More [\[+\]](#)

Rules & Requirements

Prerequisites: Master of Information and Data Science students only. DATASCI W205, DATASCI W207. Intermediate programming skills in an object-oriented language (e.g., Python)

Hours & Format

Fall and/or spring: 15 weeks - 3 hours of web-based lecture per week

Summer: 15 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Machine Learning at Scale: Read Less [\[-\]](#)

DATASCI W266 Natural Language Processing with Deep Learning 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

Understanding language is fundamental to human interaction. Our brains have evolved language-specific circuitry that helps us learn it very quickly; however, this also means that we have great difficulty explaining how exactly meaning arises from sounds and symbols. This course is a broad introduction to linguistic phenomena and our attempts to analyze them with machine learning. We will cover a wide range of concepts with a focus on practical applications such as information extraction, machine translation, sentiment analysis, and summarization.

Natural Language Processing with Deep Learning: Read More [\[+\]](#)

Rules & Requirements

Prerequisites: Master of Information and Data Science students only. Data Science W207

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Gillick

Natural Language Processing with Deep Learning: Read Less [\[-\]](#)

DATASCI W271 Statistical Methods for Discrete Response, Time Series, and Panel Data 3 Units

Terms offered: Fall 2020, Summer 2020, Spring 2020

A continuation of Data Science W203 (Exploring and Analyzing Data), this course trains data science students to apply more advanced methods from regression analysis and time series models. Central topics include linear regression, causal inference, identification strategies, and a wide-range of time series models that are frequently used by industry professionals. Throughout the course, we emphasize choosing, applying, and implementing statistical techniques to capture key patterns and generate insight from data. Students who successfully complete this course will be able to distinguish between appropriate and inappropriate techniques given the problem under consideration, the data available, and the given timeframe.

Statistical Methods for Discrete Response, Time Series, and Panel Data: Read More [\[+\]](#)

Rules & Requirements

Prerequisites: DATASCI W203 taken in Fall 2016 or later and completed with a grade of B+ or above; strong familiarity with classical linear regression modeling; strong hands-on experience in R; working knowledge of calculus and linear algebra; familiarity with differential calculus, integral calculus and matrix notations; or instructor approval

Hours & Format

Fall and/or spring: 14 weeks - 3 hours of web-based lecture per week

Summer: 14 weeks - 3 hours of web-based lecture per week

Online: This is an online course.

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Statistical Methods for Discrete Response, Time Series, and Panel Data: Read Less [\[-\]](#)