Information and Data Science: MIDS

The (https://datascience.berkeley.edu/form/)**Master of Information and Data Science (MIDS)** is an online, part-time professional degree program that prepares students to work effectively with heterogeneous, real-world data (ranging from tweet streams and call records to mouse clicks and GPS coordinates) and to extract insights from the data using the latest tools and analytical methods. The program emphasizes the importance of asking good research or business questions as well as the ethical and legal requirements of data privacy and security.

The **Fifth Year Master of Information and Data Science (Fifth Year MIDS)** is a full-time variant of the MIDS program that is designed for students who have just completed their undergraduate education at UC Berkeley.

Students attend weekly live ("synchronous") sessions with classmates and instructors via an online platform as well as engaging with online ("asynchronous") videos and assignments on their own time.

The curriculum includes research design and applications for data and analysis, storing and retrieving data, exploring and analyzing data, identifying patterns in data, and effectively visualizing and communicating data. MIDS features a project-based approach to learning and encourages the pragmatic application of a variety of different tools and methods to solve complex problems.

Graduates of the program will be able to:

- Imagine new and valuable uses for large datasets;
- Retrieve, organize, combine, clean, and store data from multiple sources;
- Apply appropriate data mining, statistical analysis, and machine learning techniques to detect patterns and make predictions;
- Design visualizations and effectively communicate findings; and
- Understand the ethical and legal requirements of data privacy and security.

The I School also offers a master's in Information Management and Systems (MIMS) (http://guide.berkeley.edu/archive/2021-22/graduate/degree-programs/information-management-systems/), a master's in Information and Cybersecurity (MICS) (http://guide.berkeley.edu/archive/2021-22/graduate/degree-programs/information-cybersecurity/), and a Ph.D (http://guide.berkeley.edu/archive/2021-22/graduate/degree-programs/information-management-systems-phd/). (http://guide.berkeley.edu/archive/2021-22/graduate/degree-programs/information-management-systems/#doctoraldegreerequirementstext)

Admission to the University

Minimum Requirements for Admission

The following minimum requirements apply to all graduate programs and will be verified by the Graduate Division:

1. A bachelor's degree or recognized equivalent from an accredited institution;
2. A grade point average of B or better (3.0);
3. If the applicant has completed a basic degree from a country or political entity (e.g., Quebec) where English is not the official language, adequate proficiency in English to do graduate work, as evidenced by a TOEFL score of at least 90 on the iBT test, 570 on the paper-and-pencil test, or an IELTS Band score of at least 7 on a 9-point scale (note that individual programs may set higher levels for any of these); and
4. Sufficient undergraduate training to do graduate work in the given field.

Applicants Who Already Hold a Graduate Degree

The Graduate Council views academic degrees not as vocational training certificates, but as evidence of broad training in research methods, independent study, and articulation of learning. Therefore, applicants who already have academic graduate degrees should be able to pursue new subject matter at an advanced level without the need to enroll in a related or similar graduate program.

Programs may consider students for an additional academic master's or professional master's degree only if the additional degree is in a distinctly different field.

Applicants admitted to a doctoral program that requires a master's degree to be earned at Berkeley as a prerequisite (even though the applicant already has a master's degree from another institution in the same or a closely allied field of study) will be permitted to undertake the second master's degree, despite the overlap in field.

The Graduate Division will admit students for a second doctoral degree only if they meet the following guidelines:

1. Applicants with doctoral degrees may be admitted for an additional doctoral degree only if that degree program is in a general area of knowledge distinctly different from the field in which they earned their original degree. For example, a physics PhD could be admitted to a doctoral degree program in music or history; however, a student with a doctoral degree in mathematics would not be permitted to add a PhD in statistics.
2. Applicants who hold the PhD degree may be admitted to a professional doctorate or professional master's degree program if there is no duplication of training involved.

Applicants may apply only to one single degree program or one concurrent degree program per admission cycle.

Required Documents for Applications

1. **Transcripts:** Applicants may upload *unofficial* transcripts with your application for the departmental initial review. *If the applicant is admitted,* then *official* transcripts of all college-level work will be required. Official transcripts must be in sealed envelopes as issued by the school(s) attended. If you have attended Berkeley, upload your unofficial transcript with your application for the departmental initial review. *If you are admitted,* an official transcript with evidence of degree conferral *will not* be required.
2. **Letters of recommendation:** Applicants may request online letters of recommendation through the online application system. Hard copies of recommendation letters must be sent directly to the program, not the Graduate Division.
3. **Evidence of English language proficiency:** All applicants who have completed a basic degree from a country or political entity in which the official language is not English are required to submit official evidence of English language proficiency. This applies to institutions from Bangladesh, Burma, Nepal, India, Pakistan, Latin America, the Middle East, the People's Republic of China, Taiwan, Japan,

Korea, Southeast Asia, most European countries, and Quebec (Canada). However, applicants who, at the time of application, have already completed at least one year of full-time academic course work with grades of B or better at a US university may submit an official transcript from the US university to fulfill this requirement. The following courses will not fulfill this requirement:

- courses in English as a Second Language,
- courses conducted in a language other than English,
- courses that will be completed after the application is submitted, and
- courses of a non-academic nature.

If applicants have previously been denied admission to Berkeley on the basis of their English language proficiency, they must submit new test scores that meet the current minimum from one of the standardized tests. Official TOEFL score reports must be sent directly from Educational Test Services (ETS). The institution code for Berkeley is 4833. Official IELTS score reports must be sent electronically from the testing center to University of California, Berkeley, Graduate Division, Sproul Hall, Rm 318 MC 5900, Berkeley, CA 94720. TOEFL and IELTS score reports are only valid for two years.

Where to Apply

Visit the Berkeley Graduate Division application page (http://grad.berkeley.edu/admissions/apply/).

Admission to the Program

Applications are evaluated holistically on a combination of prior academic performance, work experience, statement of purpose, and letters of recommendation.

The UC Berkeley School of Information seeks students with the academic abilities to meet the demands of a rigorous graduate program.

To be eligible to apply to the **Master of Information and Data Science** program, applicants must meet the following requirements:

- A bachelor's degree or its recognized equivalent from an accredited institution.
- Superior scholastic record, normally well above a 3.0 GPA.
- A high level of quantitative ability as conveyed by significant work experience that demonstrates your quantitative abilities, academic coursework that demonstrates quantitative aptitude, or scores in the top 15 percent in the Quantitative section of either the GRE or GMAT.
- A high level of analytical reasoning ability and a problem-solving mindset as demonstrated in academic and/or professional performance.
- A working knowledge of fundamental concepts including: data structures, algorithms and analysis of algorithms, and linear algebra.
- Programming proficiency as demonstrated by prior work experience or advanced coursework. (For example: Python, Java, or R.)
- The ability to communicate effectively, as demonstrated by academic performance, professional experience, strong essays that demonstrate effective communication skills, or strong scores in the Verbal and Writing sections of either the GRE or GMAT.
- A Statement of Purpose that clearly indicates professional career goals and reasons for seeking the degree.
- (Optional) Official Graduate Record Examination (GRE) (http://www.princetonreview.com/mids/) General Test or Graduate Management Admission Test (GMAT) (http://www.princetonreview.com/mids/) scores.
- Official Test of English as a Foreign Language (TOEFL) (http://www.toefl.org/) scores for applicants whose academic work has been in a country other than the US, UK, Australia, or English-speaking Canada.

Note: Admission to the **Fifth Year Master of Information and Data Science** program requires that the applicant complete their undergraduate education at UC Berkeley in the year prior to starting the program. Consequently, applicants are not required to submit GRE or GMAT scores. However, applicants are required to submit three letters of recommendation and additional short answer essays.

For more information and application instructions, please visit the datascience@berkeley Admissions Overview (http://datascience.berkeley.edu/admissions/admissions-overview/).

Unit Requirements

The Master of Information and Data Science is designed to be completed in 20 months, but other options are available to complete the program. You will complete 27 units of course work over an average of five terms, taking a maximum of 9 units each term. Courses are divided into foundation courses (15 units), advanced courses (9 units), and a synthetic capstone (3 units). You will also complete an immersion at the UC Berkeley campus.

The unit and coursework requirements for the Fifth Year Master of Information and Data Science are identical. However, as the program is full-time, it is intended to be completed in 12 months or three terms.

Curriculum

Foundation Courses		
DATASCI W200	Course Not Available	3
DATASCI W201	Course Not Available	3
DATASCI W203	Course Not Available	3
DATASCI W205	Course Not Available	3
DATASCI W207	Course Not Available	3
Advanced Courses		
DATASCI W209	Course Not Available	3
DATASCI W231	Course Not Available	3
DATASCI W233	Course Not Available	3
Not Available to Fifth Year MIDS		
DATASCI W241	Course Not Available	3
DATASCI W251	Course Not Available	3
Not Available to Fifth Year MIDS		
DATASCI W261	Course Not Available	3
Not Available to Fifth Year MIDS		
DATASCI W266	Course Not Available	3
DATASCI W271	Course Not Available	3
Not Available to Fifth Year MIDS		
Capstone Course		
DATASCI W210	Course Not Available	3

Immersion

As a Master of Information and Data Science (MIDS) student, the immersion is your opportunity to meet faculty and peers in person on the UC Berkeley campus. You will have the opportunity to gain on-the-ground perspectives from faculty and industry leaders, meet with data science professionals, and soak up more of the School of Information (I School) culture. Offered three times a year, each three- to four-day immersion will be custom-crafted to deliver additional learning, networking, and community-building opportunities.

Please refer to the datascience@berkeley website (http://datascience.berkeley.edu/academics/curriculum/) for more information.

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

Information and Data Science

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

DATASCI 200 Introduction to Data Science Programming 3 Units

Terms offered: Not yet offered

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 lecture per week

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Laskowski

Formerly known as: Data Science W200

Introduction to Data Science Programming: Read Less [-]

DATASCI 201 Research Design and Applications for Data and Analysis 3 Units

Terms offered: Not yet offered
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: MIDS students only

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Rivera

Formerly known as: Data Science W201

Research Design and Applications for Data and Analysis: Read Less [-]

DATASCI 201A Research Design and Applications for Data and Analysis for Early Career Data Scientists 4 Units

Terms offered: Fall 2022
Introduces the data sciences landscape, with a focus on learning data science techniques to uncover and answer questions students will encounter in industry. Lectures, readings, discussions, and assignments will teach how to apply methods to ask better questions, gather data, interpret results, and convey findings to various audiences. The emphasis is on making practical contributions to real decisions that organizations make. This 4-credit version of the course is designed for early-career learners in our 5th Year MIDS pathway. It provides additional attention to introducing professional and business knowledge and skills, and providing students with opportunities to apply and reflect on learning how to become a data science professional.
Research Design and Applications for Data and Analysis for Early Career Data Scientists: Read More [+]

Objectives & Outcomes

Student Learning Outcomes: By the completion of this course, students will be able to Apply techniques and approaches focused on building work relationships and engaging in interactions that align with organizational goals.
By the completion of this course, students will be able to Assess and select data and the data collection methods that best fit a specific outcome or need.

By the completion of this course, students will be able to Demonstrate an understanding of foundational approaches to project management and strategic thinking by imagining, planning, and designing a data science project from start to finish.

By the completion of this course, students will be able to Devise effective research questions and apply them to analytic processes that lead to actionable insight and strategic decisions.

By the completion of this course, students will be able to Identify and describe effective teamwork skills, practices, and characteristics of an effective workplace or project team.
By the completion of this course, students will be able to Justify and defend an analytical approach—descriptive, predictive, or explanatory—to inform efficient decision making.
By the completion of this course, students will be able to Understand and apply successful communication strategies and methods (written, spoken, and visual) for teams and for various stakeholders within an organization with different contextual requirements and expectations, including summarizing and presenting key ideas effectively for various stakeholders.

By the completion of this course, students will be able to Understand key principles that affect human decision-making processes, such as biases and contextual concerns (e.g., ethical and legal) that affect human decision-making processes and apply knowledge of those principles throughout the course to mitigate biases, facilitate better decision making, and improve communication.
By the completion of this course, students will be able to describe the role that data science as a domain and as a set of practices and processes plays in decision making made by people in organizations, and establish an awareness of common social structures, practices, norms and expectations in data science organizations, teams, and workplaces.

Rules & Requirements

Prerequisites: 5th Year MIDS Students only

Hours & Format

DATASCI 203 Statistics for Data Science 3 Units

Terms offered: Not yet offered

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: MIDS 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 lecture per week

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Formerly known as: Data Science W203

Statistics for Data Science: Read Less [-]

DATASCI 205 Fundamentals of Data Engineering 3 Units

Terms offered: Not yet offered

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: MIDS students only. Intermediate competency in Python, C, or Java, and competency in Linux, GitHub, and relevant Python libraries. Knowledge of database management including SQL is recommended but not required

Credit Restrictions: Students will receive no credit for DATASCI W205 after completing DATASCI 205. A deficient grade in DATASCI W205 may be removed by taking DATASCI 205.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Crook

Formerly known as: Data Science W205

Fundamentals of Data Engineering: Read Less [-]

DATASCI 207 Applied Machine Learning 3 Units

Terms offered: Not yet offered
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: MIDS students only. DATASCI 201 and DATASCI 203. Intermediate competency in Python, C, or Java, and competency in Linux, GitHub, and relevant Python libraries. Linear algebra is recommended

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Gillick

Formerly known as: Data Science W207

Applied Machine Learning: Read Less [-]

DATASCI 209 Data Visualization 3 Units

Terms offered: Not yet offered
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: MIDS students only. DATASCI 203. Students must take DATASCI 205 concurrently or prior to DATASCI 209. If taken concurrently, students may not drop 205 and remain in 209. 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 lecture per week

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Formerly known as: Data Science W209

Data Visualization: Read Less [-]

DATASCI 210 Capstone 3 Units

Terms offered: Not yet offered

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: MIDS students only. Must be taken in final term of the MIDS program

Credit Restrictions: Students will receive no credit for DATASCI W210 after completing DATASCI 210. A deficient grade in DATASCI W210 may be removed by taking DATASCI 210.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Formerly known as: Data Science W210

Capstone: Read Less [-]

DATASCI 231 Behind the Data: Humans and Values 3 Units

Terms offered: Not yet offered

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

Credit Restrictions: Students will receive no credit for DATASCI W231 after completing DATASCI 231. A deficient grade in DATASCI W231 may be removed by taking DATASCI 231.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Morgan

Formerly known as: Data Science W231

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

DATASCI 233 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

Credit Restrictions: Students will receive no credit for DATASCI W233 after completing DATASCI 233. A deficient grade in DATASCI W233 may be removed by taking DATASCI 233.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Formerly known as: Data Science W233

Privacy Engineering: Read Less [-]

DATASCI 241 Experiments and Causal Inference 3 Units

Terms offered: Not yet offered

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: MIDS students only. DATASCI 201 and DATASCI 203

Credit Restrictions: Students will receive no credit for DATASCI W241 after completing DATASCI 241. A deficient grade in DATASCI W241 may be removed by taking DATASCI 241.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Formerly known as: Data Science W241

Experiments and Causal Inference: Read Less [-]

DATASCI 251 Deep Learning in the Cloud and at the Edge 3 Units

Terms offered: Not yet offered
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: MIDS students only. DATASCI 201, DATASCI 203, and DATASCI 205. Students 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

Credit Restrictions: Students will receive no credit for DATASCI W251 after completing DATASCI 251. A deficient grade in DATASCI W251 may be removed by taking DATASCI 251.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Formerly known as: Data Science W251

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

DATASCI 255 Machine Learning Systems Engineering 3 Units

Terms offered: Fall 2022, Summer 2022, Spring 2022
This course provides learners hands-on data management and systems engineering experience using containers, cloud, and Kubernetes ecosystems based on current industry practice. The course will be project-based with an emphasis on how production systems are used at leading technology-focused companies and organizations. During the course, learners will build a body of knowledge around data management, architectural design, developing batch and streaming data pipelines, scheduling, and security around data including access management and auditability. We'll also cover how these tools are changing the technology landscape.
Machine Learning Systems Engineering: Read More [+]

Objectives & Outcomes

Student Learning Outcomes: Construct, measure, and identify metrics relating to performance of a system in order to optimize costs and latency of serving inferences for machine learning models.

Demonstrate understanding of Kubernetes for management of machine learning models.

Describe the difference between a monolithic and microservice architecture, assess and select appropriate use cases for each.

Describe the differences between a development and production system particularly for Machine Learning where the boundaries are blurry.

Know when to leverage a cache for serving machine learning models to reduce load on production systems.

Understand continuous integration and continuous delivery (CI/CD) pipeline for automated code deployment, particularly for ML models.

Understand how stateful systems add complexities to systems engineering.
Understand how to serve machine learning models over an API in real-time.

Rules & Requirements

Prerequisites: MIDS students only. DATASCI 205 and DATASCI 207. We assume you are familiar with generating predictions from a trained machine learning model. Familiarity with command line (Bash), Python, and Git. We assume you have a working knowledge of SSH, Ports, and familiarity with networking concepts such as DNS

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Machine Learning Systems Engineering: Read Less [-]

DATASCI 261 Machine Learning at Scale 3 Units

Terms offered: Not yet offered
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: MIDS students only. DATASCI 205 and DATASCI 207. Intermediate programming skills in an object-oriented language (e.g., Python)

Credit Restrictions: Students will receive no credit for DATASCI W261 after completing DATASCI 261. A deficient grade in DATASCI W261 may be removed by taking DATASCI 261.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Formerly known as: Data Science W261

Machine Learning at Scale: Read Less [-]

DATASCI 266 Natural Language Processing with Deep Learning 3 Units

Terms offered: Not yet offered
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: MIDS students only. DATASCI 207

Credit Restrictions: Students will receive no credit for DATASCI W266 after completing DATASCI 266. A deficient grade in DATASCI W266 may be removed by taking DATASCI 266.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Instructor: Gillick

Formerly known as: Data Science W266

Natural Language Processing with Deep Learning: Read Less [-]

DATASCI 271 Statistical Methods for Discrete Response, Time Series, and Panel Data 3 Units

Terms offered: Not yet offered

A continuation of DATASCI 203, 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: MIDS students only. DATASCI 203 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

Credit Restrictions: Students will receive no credit for DATASCI W271 after completing DATASCI 271. A deficient grade in DATASCI W271 may be removed by taking DATASCI 271.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Formerly known as: Data Science W271

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

DATASCI 281 Computer Vision 3 Units

Terms offered: Summer 2022, Spring 2022

This course introduces the theoretical and practical aspects of computer vision, covering both classical and state of the art deep-learning based approaches. This course covers everything from the basics of the image formation process in digital cameras and biological systems, through a mathematical and practical treatment of basic image processing, space/frequency representations, classical computer vision techniques for making 3-D measurements from images, and modern deep-learning based techniques for image classification and recognition.

Computer Vision: Read More [+]

Objectives & Outcomes

Student Learning Outcomes: Be able to read and understand research papers in the computer-vision literature.

Build computer vision systems to solve real-world problems.

Properly formulate problems with the appropriate mathematical and computational tools.

Understand the building blocks of classical computer vision techniques. Understand the building blocks of modern computer vision techniques (primarily artificial neural networks).

Understand the process by which images are formed and represented.

Rules & Requirements

Prerequisites: MIDS students only. DATASCI 207. We assume you are familiar with machine learning techniques. You should also be comfortable with linear algebra, which we'll use for vector representations and when we discuss deep learning. Intermediate programming skills in an object-oriented language (e.g., Python). This course will use Python for all examples, exercises, and assignments

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Computer Vision: Read Less [-]

DATASCI 290 Special Topics 3 Units

Terms offered: Fall 2022, Fall 2021
Specific topics, may vary from section to section, year to year.
Special Topics: Read More [+]
Rules & Requirements

Prerequisites: MIDS students only

Repeat rules: Course may be repeated for credit when topic changes. Students may enroll in multiple sections of this course within the same semester.

Hours & Format

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

Summer: 14 weeks - 3 hours of lecture per week

Additional Details

Subject/Course Level: Data Science/Graduate

Grading: Letter grade.

Special Topics: Read Less [-]