Information and Data Science: MIDS

The Master of Information and Data Science (MIDS) is a 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 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.

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 comes 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 (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 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 from countries or political entities in which the official language is not English are required to submit official evidence of English language proficiency. This applies to applicants 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.

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, GRE/GMAT score, 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.
- Official Graduate Record Examination (GRE) (http://www.princetonreview.com/mids) General Test or Graduate Management Admission Test (GMAT) (http://www.princetonreview.com/mids) scores.
- A high level of quantitative ability as demonstrated by scores in the top 15 percent in the Quantitative section of either the GRE or GMAT, five years of technical work experience, or significant work experience that demonstrates your quantitative abilities.
- 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 strong scores in the Verbal and Writing sections of either the GRE or GMAT, academic performance, or professional experience.
- A Statement of Purpose that clearly indicates professional career goals and reasons for seeking the degree.
- 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.

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.

Curriculum

Foundation Courses		
DATASCI W201	Research Design and Applications for Data and Analysis	3
DATASCI W203	Statistics for Data Science	3
DATASCI W205	Storing and Retrieving Data	3
DATASCI W207	Applied Machine Learning	3
DATASCI W209	Data Visualization and Communication	3
Advanced Courses		
DATASCI W231	Behind the Data: Humans and Values	3
DATASCI W241	Experiments and Causal Inference	3
DATASCI W251	Scaling Up! Really Big Data	3
DATASCI W261	Machine Learning at Scale	3
DATASCI W271	Statistical Methods for Discrete Response, Time Series, and Panel Data	3
Capstone Course		
DATASCI W210	Capstone	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 twice a year, each four- to five-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.

Information and Data Science

DATASCI W200 Python Fundamentals for Data Science 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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.
Python Fundamentals for Data Science: 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: 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.

Instructor: Laskowski

Python Fundamentals for Data Science: Read Less [-]

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

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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: 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.

Instructor: Weber

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

DATASCI W203 Statistics for Data Science 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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: 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.

Instructor: Cheshire

Statistics for Data Science: Read Less [-]

DATASCI W205 Storing and Retrieving Data 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
Data Science depends on data, and a core competency mandated by this reliance on data is knowing effective and efficient ways to manage, search and compute over that data. This course is focused on how data can be stored, managed and retrieved as needed for use in analysis or operations. The goal of this course is provide students with both theoretical knowledge and practical experience leading to mastery of data management, storage and retrieval with very large-scale data sets. Course must be taken for a letter grade to fulfill degree requirements.
Storing and Retrieving Data: 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.

Instructor: Larson

Storing and Retrieving Data: Read Less [-]

DATASCI W207 Applied Machine Learning 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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: 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.

Instructor: Hearst

Applied Machine Learning: Read Less [-]

DATASCI W209 Data Visualization and Communication 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
Communicating clearly and effectively about the patterns we find in data is a key skill for a successful data scientist. This course focuses on the design and implementation of complementary visual and verbal representations of patterns and analyses in order to convey findings, answer questions, drive decisions, and provide persuasive evidence supported by data. Assignments will give hands-on experience designing data graphics and visualizations, and reporting findings in prose. Course must be taken for a letter grade to fulfill degree requirements.
Data Visualization and Communication: Read More [+]

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

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.

Instructor: Ryokai

Data Visualization and Communication: Read Less [-]

DATASCI W210 Capstone 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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: Master of Information and Data Science Students only.
Students must have completed (or are completing during the same semester) all core courses (Data Science W201, W203, W205, W207 and W209)

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.

Capstone: Read Less [-]

DATASCI W231 Behind the Data: Humans and Values 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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: 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.

Instructor: Mulligan

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

DATASCI W241 Experiments and Causal Inference 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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 Scaling Up! Really Big Data 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
An overview of the contemporary toolkits for problems related to cloud computing and big data. Because the class is an advanced course, we generally assume familiarity with the concepts and spend more time on the implementation. Every lecture is followed by a hands-on assignment, where students get to experience some of the technologies covered in the lecture. By the time students complete the course, they should be able to name the big data problem they are facing, select proper tooling, and know enough to start applying it.
Scaling Up! Really Big Data: 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: 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.

Scaling Up! Really Big Data: Read Less [-]

DATASCI W261 Machine Learning at Scale 3 Units

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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.
Data Science 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

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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: 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.

Instructor: Daniel Gillick

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

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

Offered through: Information
Terms offered: Spring 2018, Fall 2017, Summer 2017
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 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: 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.

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