Master's in Data Science

Data Science Program

Humanity's inexorable digitization of existing data, creation of born-digital data, and the rapidly evolving Internet are driving the need for global transformation of higher education and its research, education, and service missions. It has transformed scholarship in numerous fields including the life sciences, engineering, natural sciences, social sciences, as well as in the arts and humanities. The University of Michigan Master of Science in Data Science is a professional degree equipped with strong methodological training. The degree program will require at least 25 units of coursework. The specific requirements are expressed in terms of foundational training in Computing and Statistics, and 4 core areas in which proficiency is required. Students cumulative GPA must be 3.00 (B) or better to stay in good standing.

Prerequisites

Prospective students must have completed 2 semesters of college calculus, 1 semester of linear algebra, and 1 introduction to computing course.

Courses

Students must take the following pre-core courses (unless waived through the course review process):

MATH 465: Introduction to Combinatorics
EECS 402: Programming for Scientists and Engineers
EECS 403: Data Structures for Scientists and Engineers
1 of the following
- BIOSTATS 601: Probability and Distribution
- STATS 425: Introduction to Probability
- STATS 510: Probability and Distribution

1 of the following
- BIOSTATS 602: Biostatistical Inference
- STATS 426: Introduction to Theoretical Statistics
- STATS 511: Statistical Inference

All Students must take the following core courses:

EECS 409: Data Science Colloquium

Expertise in Data Management and Manipulation
1 of the following
- EECS 484: Database Management Systems
- EECS 584: Advanced Database Systems

1 of the following
- EECS 485: Web Systems
- EECS 486: Information Retrieval and Web Search
- EECS 549/SI 650: Information Retrieval
- SI 618: Data Manipulation Analysis
- STATS 507: Data Science and Analytics using Python

Expertise in Data Science Techniques
1 of the following:
- BIOSTAT 650: Data Manipulation and Analysis
- STATS 413: Applied Regression Analysis
- STATS 513: Regression and Data Analysis
- STATS 500: Statistical Learning I: Linear Regression

1 from the following:
- BIOSTAT 616: Machine Learning for Health Sciences
- EECS 545: Machine Learning
- SI 670: Applied Machine Learning
- SI 671: Data Mining: Methods and Applications
- STATS 415: Data Mining and Statistical Learning
- STATS 503: Statistical Learning II: Multivariate Analysis

Capstone
- BIOSTAT 610: Reading in Biostatistics
- BIOSTAT 629: Case Studies for Health Big Data
- BIOSTAT 698: Modern Statistical Methods in Epidemiologic Studies
- BIOSTAT 699: Analysis of Biostatistical Investigations
- EECS 588: Directed Study
- SI 599-00X: Computational Social Science

- SI 699-004: Big Data Analytics
- STATS 504: Principles and Practices in Effective Statistical Consulting
- STATS 750: Directed Reading

Electives:

Select 1 from each competency. Students may not count a course in multiple categories. Electives group must include at least two advanced graduate courses.

Principles of Data Science

BIOSTAT 601 (Probability and Distribution Theory) | BIOSTAT 602 (Biostatistical Inference) | BIOSTAT 617 (Sample Design) | BIOSTAT 626 (Machine Learning Methods) | BIOSTAT 680 (Stochastic Processes) | BIOSTAT 682 (Bayesian Analysis) | EECS 501 (Probability and Random Processes) | EECS 502 (Stochastic Processes) | EECS 551 (Matrix Methods for Signal Processing, Data Analysis and Machine Learning) | EECS 553 (Theory and Practice of Data Compression) | EECS 564 (Estimation, Filtering, and Detection) | SI 670 (Applied Machine Learning | STATS 451 (Introduction to Bayesian Data Analysis) | STATS 470 (Introduction to Design of Experiments) | STATS 510 (Probability and Distribution Theory) | STATS 511 (Statistical Inference) | STATS 551 (Bayesian Modeling and Computation) | STATS 570 (Design of Experiments)

Data Analysis

BIOSTAT 645 (Time series) | BIOSTAT 651 (Generalized Linear Models) | BIOSTAT 653 (Longitudinal Analysis) |BIOSTAT 665 (Population Genetics) | BIOSTAT 666 (Statistical Models and Numerical Methods in Human Genetics) | BIOSTAT 675 (Survival Analysis) | BIOSTAT 685 (Non-parametric statistics) | BIOSTAT 695 (Categorical Data) | BIOSTAT 696 (Spatial statistics) | EECS 556 (Image Processing) | EECS 559 (Advanced Signal Processing) | EECS 659 (Adaptive Signal Processing) | STATS 414 (Topics in Applied Data Analysis) | STATS 449 (Applied Survival Analysis) | STATS 501 (Statistical Analysis of Correlated Data) | STATS 503 (Statistical Learning II: Multivariate Analysis) | STATS 509 (Statistics for Financial Data) | STATS 531 (Analysis of Time Series) | STATS 600 (Linear Models) | STATS 601 (Analysis of Multivariate and Categorical Data) | STATS 605 (Advanced Topics in Modeling and Data Analysis) | STATS 700 (Topics in Applied Statistics)

Computation

BIOSTAT 607 (Basic Computing in Data Analytics) | BIOSTAT 615 (Statistical Computing) | BIOSTAT 625 (Computing with Big Data) | EECS 481 (Software Engineering) | EECS 485 (Web Systems) | EECS 486 (Information Retrieval and Web Search) | EECS 490 (Programing Languages) | EECS 493 (User Interface Development) | EECS 504 (Computer Vision) |EECS 542 (Advanced Topics in Computer Vision) | EECS 548/SI 649 (Information Visualization) | EECS 549/SI 650 (Information Retrieval) | EECS 586 (Design and Analysis of Algorithms) | EECS 587 (Parallel Computing) | EECS 592 (Artificial Intelligence) | EECS 595/SI 561 (Natural Language Processing) | SI 608 (Networks) | SI 630 (Natural Language Processing (Algorithms and People) | SI 671 (Data Mining: Methods and Applications) | STATS 406 (Computational Methods in Statistics and Data Science) | STATS 506 (Computational Methods and Tools in Statistics) | STATS 507 (Data Science and Analytics using Python) | STATS 606 (Statistical Computing) | STATS 607 (Programming and Numerical Methods in Statistics) | STATS 608 (Monte Carlo Methods and Optimization Methods in Statistics)

Program notes:

• At least 25 units of graduate-level coursework.

• At least 18 units of coursework at the advanced graduate level (500 level or above in LSA, UMSI, and CoE, or 600 level or above in SPH).