**University of Michigan**

**Data Science Master's Program**

Data Science is often viewed as the confluence of (1) Computer and Information Sciences (2) Statistical Sciences, and (3) Domain Expertise. These three pillars are not symmetric: the first two together represent the core methodologies and the techniques used in Data Science, while the third pillar is the application domain to which this methodology is applied. In this program, core data science training is focused on the first two pillars, along with practice in applying their skills to address problems in application domains.

We characterize the required Data Science skills in two categories: statistical skills, such as those taught by the Statistics and Biostatistics departments, and computational skills, such as those taught by the Computer Science and Engineering Division and the School of Information. The design of the program is to require every student to receive balanced training in both areas. To create an academic plan that achieves this balance, and to foster a greater sense of shared community, we do not intend to offer any sub-plans or tracks within the proposed degree program. Rather, we will expect graduates of this program to understand data representation and analysis at an advanced level.

With the MS in Data Science all students will be able to: identify relevant datasets, apply the appropriate statistical and computational tools to the dataset to answer questions posed by individuals, organizations or governmental agencies, design and evaluate analytical procedures appropriate to the data, and implement these efficiently over large heterogeneous data sets in a multi-computer environment.

**Prerequisites**

Our diverse community of graduate students comes from many different countries and many undergraduate majors, including statistics, mathematics, computer science, physics, engineering, information, and data science. While a Data Science undergraduate major is not required, it is expected that applicants will have at least the following background before they join:

* 2 semesters of college calculus
* 1 semester of linear or matrix algebra
* 1 introduction to computing course

**Courses**

**Students must take the following core courses (unless waived by the course review process):**

MATH 403: Introduction to Discrete Mathematics

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 Analytics using Python

**Expertise in Data Science Techniques**

1 of the following:

* BIOSTAT 650: Applied Statistics I: Linear Regression
* STATS 500: Statistical Learning I: Linear Regression
* STATS 513: Regression and Data Analysis

1 from the following:

* STATS 415: Data Mining and Statistical Learning
* STATS 503: Statistical Learning II: Multivariate Analysis
* EECS 505: Computational Data Science and ML
* EECS 545: Machine Learning
* EECS 476: Data Mining
* EECS 576: Advanced Data Mining
* SI 670: Applied Machine Learning
* SI 671: Data Mining: Methods and Applications
* BIOSTAT 626: Machine Learning for Health Sciences

**Capstone**

* STATS 504: Principles and Practices in Effective Statistical Consulting
* STATS 750: Directed Reading
* EECS 599: Directed Study
* SI 599-00X: Computational Social Science
* SI 691: Independent Study
* SI 699-004: Big Data Analytics
* 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

**Electives**

Select 1 course of at least 3 credits from each group. Electives must include at least two advanced graduate courses (500 level or above in LSA, UMSI, and CoE, or 600 level or above in SPH). Students may not double-count a course in multiple categories.

*Principles of Data Science*

BIOSTAT 601 (Probability and Distribution Theory) | BIOSTAT 602 (Biostatistical Inference) | BIOSTAT 617 (Sample Design) | BIOSTAT 626 (Machine Learning for Health Sciences) | 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 559 (Optimization Methods for SIPML) | 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)

*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 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 615 (Statistical Computing) | BIOSTATS 625 (Computing with Big Data) | EECS 481 (Software Engineering) | EECS 485 (Web Systems) | EECS 486 (Information Retrieval and Web Search) | EECS 490 (Programming Langiages) | EECS 493 (User Interface Development) | EECS 504 (Computer Vision) |EECS 542 (Advanced Topics in Computer Vision) | EECS 549/SI 650 (Information Retrieval) | EECS 548/SI 649 (Information Visualization) | 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 618 (Data Manipulation and Analysis | 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 507 (Data Science Analytics using Python) | STATS 506 (Computational Methods and Tools in Statistics) | STATS 606 (Statistical Computing) | STATS 608 (Monte Carlo Methods and Optimization Methods in Statistics)

**Program Notes**

·         At least 25 units of graduate-level coursework must be completed during residency in the Data Science program. Of these 25, 18 must be at the advanced graduate level (500 level or above in LSA, UMSI, and CoE, and 600 level or above in SPH).

·         Expertise in Data Science Techniques part 1 can be fulfilled by STATS 413 if taken before program start.

·         Expertise in Data Science Techniques part 2 can be fulfilled by EECS 445 if taken before program start.

·         Program requirements: the cumulative GPA must be B (3.0) or better as required by Rackham.