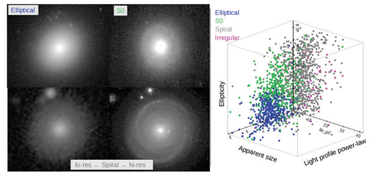
What Does a Data Scientist Do?

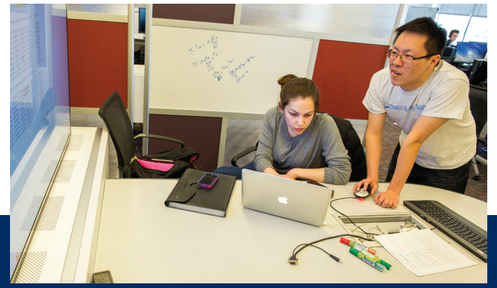
A Data Scientist Explores Data From Multiple Disparate Sources to Extract Previously Hidden Insights

Data science is a new field that combines creative, inquisitive thinking with computational and statistical techniques to explore, manage, analyze, and interpret data in a way that can influence how an organization approaches a business or research challenge. Good data scientists will not just address problems, they will pick the right approaches for the right problems that have the most value to their organization.

A data scientist will analyze mixed, unstructured, and large sets of data with the goal of discovering a previously hidden insight, which in turn can provide a competitive advantage or address a pressing problem. A data scientist does not simply collect and report on data, but also looks at it from many angles, determines what it means, and then develops informed conclusions and recommends ways to apply the data.



Low and high resolution galaxy images are collected and analyzed for size, ellipticity, and intensity in order to classify galaxy morphologies and to separate spirals from ellipticals, even in poorly resolved images. Image courtesy of Chris Miller, U-M Astronomy, Guillermo Cabrera Center for Mathematical Modeling, Univ de. Chile.



Data science involves teamwork and collaboration.

Developed by CSE and Statistics

The Data Science program was developed jointly by the Computer Science and Engineering Division in the College of Engineering and the Department of Statistics in the College of Literature, Science, and the Arts. Faculty from both departments teach the courses.

CSE Division, Bob and Betty Beyster Building

Founded by computing pioneer Arthur Burks in 1956, the Computer Science and Engineering Division is one of the oldest and most respected in the world. Students affiliated with the division study under internationally-recognized faculty in programs that emphasize the technical fundamentals of computer science, as well as computational thinking and the creative application of computing.

CSE COMPUTER SCIENCE AND ENGINEERING UNIVERSITY OF MICHIGAN

The Department of Statistics is regarded as an international leader in statistical education and research, where faculty explore some of the most interesting and relevant problems of the day. Students affiliated with Statistics engage with award-winning faculty who are developing new methods and insights into the collection, analysis, interpretation, presentation, and organization of data.

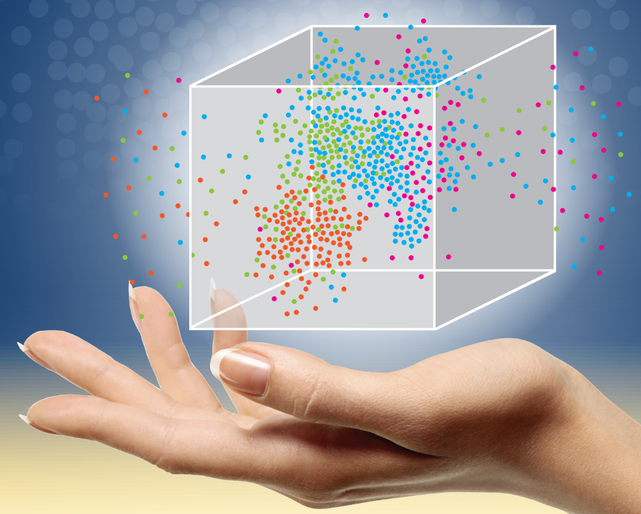
LSA COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS UNIVERSITY OF MICHIGAN

Statistics Department, West Hall

Data Science @ Michigan

Department of Electrical Engineering and Computer Science

DATA SCIENCE



cse.umich.edu

2808 Beyster Bldg. • 734-763-6563 • ugadmin@umich.edu

MICHIGAN ENGINEERING UNIVERSITY OF MICHIGAN

One of the First Programs in Data Science

Extracting Knowledge From an Ocean of Data

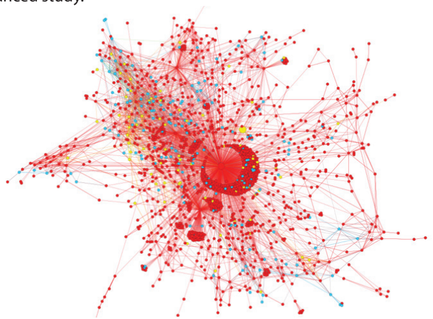
Big data, unstructured data, mixed media data, you name it – a need exists for a new class of experts who can extract actionable knowledge from rich and varied datasets.

Huge amounts of data with complex structures in the form of text, imagery, video, and streaming data are routinely collected in social networks, biological and health sciences, business and industry, as well as by government and society at large. Data scientists will collect, curate, manage, analyze, and transform this data into information that can enable knowledge creation and decision-making.

A Multidisciplinary Program that Draws From Computer Science, Statistics, and Mathematics

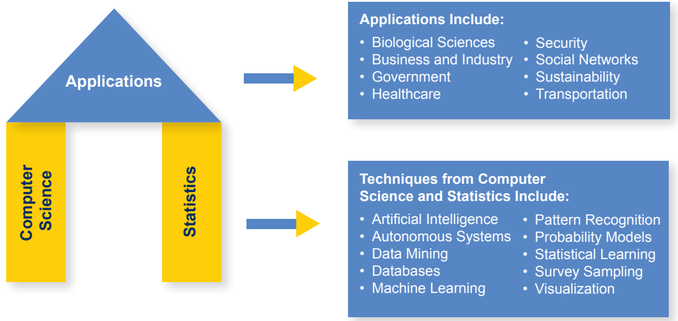
The Data Science program emphasizes fundamental methods from the fields of computer science, statistics, and mathematics, along with data analytics techniques that are essential for analyzing the growing amount of data generated in today's business, science, and engineering applications. Furthermore, the program exposes students to a variety of application domains through threads of upper-level electives, including issues of ethics, privacy, and security.

The program will emphasize both the practical use of Data Science methods as well as the theoretical properties underpinning the performance of the methods and algorithms. Major Design Experience courses in the program have been designed with input from industry leaders and will give students the opportunity to address real-world challenges. Graduates of the program will be well prepared for exciting careers in Data Science as well as for advanced study.

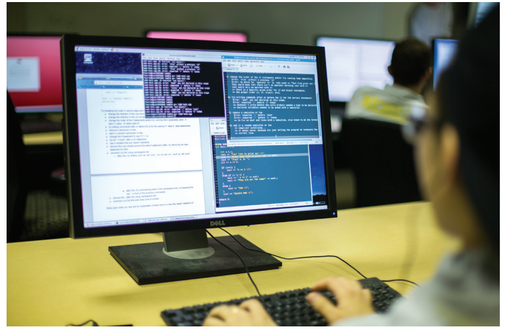


Michigan researchers have developed software to help identify and correct false claims on Twitter.

Discovery in the Age of Big Data



Data science leverages methods and techniques from computer science and statistics to enable new discoveries.



Leaders in Data Science

I like the term data scientist, because it expresses where we are today – running expeditions into overwhelming amounts of data.

Usama Fayyad (BSE '84; MSE '86, PhD '91), Chief Data Officer and Group Managing Director at Barclays Bank. Prior to Barclays, Dr. Fayyad was the first person to hold the title of Chief Data Officer at Yahoo. He previously led efforts to draw actionable conclusions from big data at NASA's Jet Propulsion Laboratory and at Microsoft.

Watson is accelerating a new class of software, services, and apps that leverage big data to think, learn, and fuel new discoveries.

Michael Rhodin (BSE '84) is Senior Vice President for Watson, one of IBM's most significant innovations in the company's 100-year history. Before heading up Watson, he led IBM's Software Solutions Group delivering industry-specific solutions in high-growth areas such as business analytics, smarter commerce, smarter cities, and social business.

Social media lives on big data. We have to look below the surface to allow our users to create their worlds.

Erin Teague (BSE '04, MBA Harvard '08), Director of Product Management at Yahoo. She is responsible for the new user experience product development and management for Yahoo's products worldwide. Prior to Yahoo, Erin worked as a Product Manager at Path and Twitter where she focused on user growth, product strategy, and analytics.

Michigan faculty are at the forefront in developing tools and techniques for Data Science.

Michael Cafarella, who is now on the faculty at Michigan, and Doug Cutting developed Hadoop, an open-source software framework for distributed storage and processing of very large data sets on computer clusters. Hadoop is an enabling tool for the work conducted by data scientists.