MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

Data Science: Engineering

WHAT IS DATA SCIENCE?

The scope of data science includes methods and algorithms involved in collecting, curating, managing, analyzing, and transforming data into information so as to enable knowledge creation and decision-making in a variety of application domains. This could include healthcare, social networking, transportation, marketing, sports/entertainment, finance, and economics. A data scientist will draw on techniques from both computer science and from statistics to help with drug discovery, analyze social network graphs, help people drive more safely and efficiently, help companies reach the right customers, and help companies or governments make better data-driven predictions. A data scientist is familiar with methods from data structures and algorithms, databases, machine learning, statistical modeling, and has exposure to one or more application domains. Data scientists are inquisitive by nature, like to question assumptions and processes, and enjoy communicating their findings from data to others.

Information from: https://www.eecs.umich.edu/eecs/undergraduate/data-science/dse_guide_fall_2015.pdf

NEEDED SKILLS:

- Computer programming skills, including familiarity with both software and hardware
- Algorithmic thinking
- Problem solving
- Communication and interpersonal skills
- Data analysis
- Mathematical and logical reasoning
- Teamwork

JOB TITLES

- Data Scientist
- Software Engineer
- Software Developer

SALARIES

$113,436

The nationwide average salary for data scientists.

**Information from: http://www.glassdoor.com on June 22nd, 2016.*

INDUSTRIES AND OCCUPATIONS

- Software industry
- Graduate school in CS, Statistics, or Analytics/Academics
- Federal Government
- Internet-based companies
- On-demand service companies
- Business consulting & management
- Financial institutions

JOB OUTLOOK

Data science is expected to be a growing field with strong demand in an Internet-based and data-driven economy. Most businesses are increasingly leveraging information to improve their processes. Graduate school opportunities are strong as there is very active research in machine learning and big data in most computer science as well as statistics departments. See the links below for a Harvard Business Review article, a report by the recruiting firm Burtch Works, and search on Glassdoor for data scientist jobs.

MORE INFORMATION

- stats.bls.gov/ooh
- Harvard Business Review: goo.gl/oP5Qhw
- Burtch Works: goo.gl/oP5Qhw
- Glassdoor: http://www.glassdoor.com
- http://www.acm.org (Association for Computing Machinery)
- http://www.computer.org/portal/siste/ieeecs/ (IEEE Computing
- Society)
- Engineering Career Resource Center, 230 Chrysler
- See a DS-Engr advisor. Sign up on the EAC website or contact the CSE Undergraduate Advising Office at 2808 Beyster Bldg. or 763-6563

MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

Data Science: Engineering

WHICH DS CLASSES SHOULD YOU START WITH?

To begin the DS-Eng major, first take EECS 203 and EECS 280, followed by EECS 281. Advanced data science courses rely on linear algebra, probability, and statistics. Thus, it is wise to take the math course sequence and STATS 412 early. STATS 412 provides a balanced exposure to both probability and statistics. **Note that STATS 250 is not adequate for the major.** If you have already taken 250 or another probability/statistics class, please see a DS-Eng advisor for advice on specific courses that you may be required to take to fill the gap. Read more about the DS-Eng major and EECS Department at: http://www.eecs.umich.edu/undergraduate/data-science

COURSE DESCRIPTIONS

EECS 203 – 4 credits
Discrete Math

Prerequisite: Math 115.
Introduction to the mathematical foundations of computer science. Topics covered include propositional and predicate logic, set theory, function and relations, growth of functions and asymptotic notation, introduction to algorithms, elementary combinatorics and graph theory, and discrete probability theory.

EECS 280 – 4 credits
Programming and Introductory Data Structures

Prerequisites: Math 115 and ENGR 101/EECS 183/ENGR 151.
Techniques and algorithm development and effective programming, top-down analysis, structured programming, testing, and program correctness. Program language syntax and static and runtime semantics. Scope, procedure instantiation, recursion, abstract data types, and parameter passing methods. Structured data types, pointers, linked data structures, stacks, queues, arrays, records, and trees.

EECS 281 – 4 credits
Data Structures and Algorithms

Prerequisites: EECS 203 and EECS 280.
Introduction to algorithm analysis and O-notation; fundamental data structures including lists, stacks, queues, priority queues, hash tables, binary trees, search trees, balanced trees and graphs, searching and sorting algorithms, recursive algorithms, basic graph algorithms, introduction to greedy algorithms and divide and conquer strategy. Several programming assignments.

STATS 412 – 3 credits
Introduction to Probability and Statistics

Prerequisite: Prior or concurrent enrollment in MATH 215.
The objectives of this course are to introduce students to the basic ideas of probability and statistical inference and to acquaint students with some important data analytic techniques, such as regression and the analysis of variance. Examples will emphasize applications to the natural sciences and engineering. There will be regular homework, two midterms, and a final exam.

EECS 445 – 4 credits
Introduction to Machine Learning

Prerequisite: EECS 281. Theory and implementation of state of the art machine learning algorithms for large-scale real-world applications. Topics include supervised learning (regression, classification, kernal methods, neural networks, and regularization) and unsupervised learning, (clustering, density estimation, and dimensionality and reduction).

STATS 413 – 4 credits
Applied Regression Analysis

Advisory prerequisite: Math 214/Math217/Math417 and {(Stats 250 and Math/Stats 425) or Stats 412 or Stats 426}.
The following topics will be covered: a) models and methods of inference for simple and multiple regression, regression splines; b) diagnostics, multicollinearity, influence, outliers, transformation, model selection, and dimension reduction; c) principal component regression, ridge and robust regression, non-linear regression, non-parametric regression, and Lasso; d) generalized linear models, binary and Poisson regression.