

Data Science

Introduction

Data Science, the science of extracting meaningful and useful information from large data sets, is a field that is developing fast. The subject is, by its inherent nature, highly interdisciplinary. There is a rapidly growing demand for data scientists in many sectors, including digital advertising, actuarial research, customized health services, genomics, image analysis, image and speech recognition, and online map and navigation services.

Prerequisites

The following courses are prerequisites for the required courses in the Data Science major:

CSE 1010 Introduction to Computing for Engineers
 MATH 1131Q Calculus I (or MATH 2141Q)
 MATH 1132Q Calculus II (or MATH 2142Q)

Required Courses

Your individualized major must include at least 36 credits of 2000-level or higher courses, 18 of which must be from the College of Liberal Arts and Sciences. The following must be part of the plan of study:

MATH 2210Q Applied Linear Algebra
STAT 3025Q Statistical Methods
STAT 3115Q Analysis of Experiments
CSE 2050 Data Structures and Object-Oriented Design
CSE 2102 Introduction to Software Engineering
CSE 2500 Introduction to Discrete Systems

UNIV 4697W Senior Thesis. A major component of this program is a three-credit, final-year capstone project, which incorporates real data science problems, requiring integration of statistics, mathematics, computing, domain knowledge, as well as soft skills such as collaboration and communication.

Writing course. A writing intensive course that is relevant to the major theme. This will normally be designated “W” in the course catalog and is *in addition to* the thesis.

Electives

In addition to the required courses above, the plan of study must include 15 credits of electives. At least six credits of these must be from a *domain science* – a scientific domain in which “big data” is used.

There are many options for domain sciences; below are three *examples*:

- For the domain of economics: ECON 2311Q Econometrics I and ECON 2312Q Econometrics II.
- For the domain of actuarial science: MATH 2620 Financial Mathematics I and MATH 3630 Long-Term Actuarial Mathematics I.
- For the domain of bioinformatics: CSE 3800 Bioinformatics, MCB 3421 Intro to Molecular Evolution and Bioinformatics.

Course List

The following is a list of courses that may be relevant to a major in Data Science. **Please note** that this is not a complete list – you may find other relevant courses in the catalog.

CSE

3100 Systems Programming
 3500 Algorithms and Complexity
 3666 Introduction to Computer Architecture
 3800 Bioinformatics
 3802 Numerical Methods in Scientific Computation
 4300 Operating Systems
 4502 Big Data Analytics
 4701 Principles of Databases
 4705 Artificial Intelligence
 4820 Intro to Machine Learning

EEB

4100 Big Data Science for Biologists

ECON

2311Q Econometrics I
 2312Q Econometrics II
 2326 Operations Research
 3313 Elementary Economic Forecasting
 3315 Financial Econometrics

Both or one of the following are prerequisites for 3000-level courses and are also useful for students with economics as a domain science:
 2211Q or 2201 Intermediate Microeconomics
 2212Q or 2202 Intermediate Macroeconomics

For qualified students there is additional relevant coursework at the graduate level.

MATH

2620 Financial Mathematics I
 3160 Probability
 3370 Differential Geometry
 3510 Numerical Analysis I
 3620 Foundations of Actuarial Science
 3630 Long-Term Actuarial Mathematics I

MCB

3421 Introduction to Molecular Evolution and Bioinformatics

For qualified students there is additional relevant coursework at the graduate level.

OPIM

3103 Business Information Systems
 3802 Data and Text Mining
 3804 Data Visualization

STAT

3375Q Introduction to Mathematical Statistics I
 3445 Introduction to Mathematical Statistics II
 3515Q Design of Experiments
 3675Q Statistical Computing
 4185 Special Topics: Intro to Data Science
 4625 Intro to Biostatistics
 4825 Applied Time Series