DATA SCIENCE (DATA)

undefined

DATA Courses

DATA 100. Data Science for All I. 4 units
Term Typically Offered: F, W, SP
2020-21 or later catalog: GE Area B4
2019-20 or earlier catalog: GE Area B4
Prerequisite: MATH 115, MATH 116, MATH 118, or Appropriate Math Placement Level.

Basic approaches for answering questions using data. Emphasis on working with tabular data in spreadsheet software to provide insights via descriptions and visualizations. Methods of acquiring data. Sampling bias, variability, and multi-variable thinking. Introduction to data modeling and data ethics. Intended for students in non-computing disciplines. Not open to students with credit in STAT 150, STAT 252, STAT 302, STAT 312, or STAT 313. Course may be offered in a classroom-based, online, or hybrid format. 4 lectures. Fulfills GE Area B4 (GE Area B1 for students on the 2019-20 or earlier catalogs); a grade of C- or better is required in one course in this GE area.

DATA 301. Introduction to Data Science. 4 units
Term Typically Offered: F, W, SP
Prerequisite: CPE/CSC 202; and one of the following: IME 326, STAT 302, STAT 312, or STAT 313.

Introduction to the field of data science and the workflow of a data scientist. Types of data (tabular, textual, sparse, structured, temporal, geospatial), basic data management and manipulation, simple summaries, and visualization. Course may be offered in classroom-based, online, or hybrid format. 3 lectures, 1 laboratory.

DATA 401. Data Science Process and Ethics. 3 units
Term Typically Offered: F
Prerequisites: DATA 301; CSC 365; and CSC 466. Concurrent: DATA 402 and DATA 403.

Complete life cycle of a data science project. Requirements engineering and data acquisition. Management and integration of data of high volume, velocity, and variety. Deployment of data science products. Engagement with stakeholders. Ethical considerations, including privacy and fairness. 3 lectures.

DATA 402. Mathematical Foundations of Data Science. 3 units
Term Typically Offered: F
Prerequisites: CSC 466; DATA 301; and STAT 334. Concurrent: DATA 401 and DATA 403.

Mathematical foundations of machine learning and data science. Principle of maximum likelihood. Inferential and predictive modeling and their comparison. Optimization techniques. Linear regression and linear classifiers. Mathematical foundations of neural networks and neural network analysis. Dimensionality reduction and its use in supervised and unsupervised learning. 3 lectures.

DATA 403. Data Science Projects Laboratory. 1 unit
Term Typically Offered: F
Concurrent: DATA 401 and DATA 402.

Project-based lab component of DATA 401 and DATA 402. Projects involving comparison of predictive and interpretable regression models, implementing linear classifiers with gradient descent, implementing neural networks from scratch, and deep learning. 1 laboratory.

DATA 441. Bioinformatics Capstone I. 2 units
Term Typically Offered: W
Prerequisite: BIO 351 or CHEM 373; BIO 441 or CSC 448; DATA 301.

Working with clients to design bioinformatics solutions to biological questions. Software requirements, elicitation techniques, data gathering, project planning, and project team organization. Ethics and professionalism. 2 laboratories.

DATA 442. Bioinformatics Capstone II. 2 units
Term Typically Offered: SP
Prerequisite: DATA 441.

Continue projects initiated in DATA 441. Team-based design, implementation of bioinformatics solutions and management of development teams. Technical documentation, quality assurance, and systems testing. Design and conduct empirical studies. Data visualization. Oral and written presentation. 2 laboratories.

DATA 451. Data Science Capstone I. 2 units
Term Typically Offered: W
Prerequisite: DATA 401.

Working with clients to develop data-driven solutions for systems to be constructed in DATA 452. Specification and design requirements, elicitation techniques, research and data gathering methods; project planning, time and budget estimating; project team organization. Ethics and professionalism. 2 laboratories.

DATA 452. Data Science Capstone II. 2 units
Term Typically Offered: SP
Prerequisite: DATA 451.

Team-based design, implementation, deployment and delivery of a system or analytical methodology that involves working with and analyzing large quantities of data. Technical management of research and development teams. Technical documentation, quality assurance, integration and systems testing. Design and conduct of empirical studies. Visualization and presentation of results orally and in writing. 2 laboratories.

DATA 472. Data Science Seminar. 1 unit
Term Typically Offered: F, W, SP
CR/NC
Prerequisite: DATA 301.

Discussions of technical, societal and ethical aspects of modern data science theory and practice, concentrating on topics not covered in other courses. 1 seminar. Total credit limited to 4 units. Credit/No credit grading only.