

Data Science and Engineering

VISION

In the Data Science and Engineering program, engineering professionals combine the skills of software programmer, database manager, and statistician to create mathematical models of the data, identify trends/deviations, and then present them in effective visual ways that can be understood by others.

Data scientists unlock new sources of economic value, provide fresh insights into science, and inform decision makers by analyzing large, diverse, complex, longitudinal, and distributed data sets generated from instruments, sensors, internet transactions, email, video, and other digital sources.

Students entering the MAS program for a degree in Data Science and Engineering will undertake courses in programming, analysis, and applications management and visualization. This program requires three foundational courses, four core courses, and two electives totaling thirty-four units, plus a capstone team project course of four units, for a total of thirty-eight units.

DATA SCIENTISTS UNDERSTAND SYSTEMS THAT CAN BE USED TO PREDICT OR IDENTIFY MAJOR EVENTS BY MODELING SYSTEMS WHICH MANIFEST THEMSELVES THROUGH DATA THAT IS OFTEN MUCH LARGER THAN THE CAPACITY OF TYPICAL MACHINES

EARN A GRADUATE DEGREE FROM A WORLD CLASS INSTITUTION AND FACULTY IN A UNIQUE, CROSS-DISCIPLINARY FIELD.

Emergence of a technical leader in engineering design that will enhance career experiences and help to prepare for future technical challenges.

Courses taught by UC San Diego's world-class faculty from the Computer Science & Engineering department and the San Diego Supercomputer Center.

Earn a master's degree in two years. Courses conveniently offered Friday/Saturday.

Data Science and Engineering Program

Master of Advanced Study Degree

ABOUT THE MASTER OF ADVANCED STUDY DEGREE

The Master of Advanced Study (MAS) is a unique multidisciplinary degree program focused on emerging technology areas that will be crucial for the future of engineering advancements. Courses will be taught by faculty in the Computer Science and Engineering department and the San Diego Supercomputer Center.

This high quality degree program is offered to working engineering professionals in industry or government. Courses will be conveniently delivered on alternating weekends, on a Friday/Saturday schedule, with instructional materials available online.

WHO SHOULD APPLY

The program is targeted towards engineering professionals working in industries concerned with managing and utilizing large data sets.

The program is designed for early to mid-career data scientists who are on a technical leadership track within their companies, or who are interested in expanding their knowledge and training in data science.

HOW TO APPLY

Visit JacobsSchool.ucsd.edu/MAS/DSE/ for complete application procedures.

FACULTY DIRECTORS



Yoav Freund
Professor
Computer Science and Engineering



Alin Deutsch
Professor
Computer Science and Engineering



Ilkay Altintas
Chief Data Science Officer
San Diego Supercomputer Center

COURSEWORK

Fall Quarter-Year One

Python for Data Analysis (4 units)

Introduction to object oriented programming using Python. Regular expressions. NumPy and Numerical Processing. IPython and plotting; data analysis using PANDAS. Webpage scraping using Scrapy; Twitter API. NLTK.

Case Studies in Data Science (2 units)

Case studies discussed by speakers from industry, government and academia expose students to the needs and uses of different technologies and their roles in model building.

Winter Quarter-Year One

Data Management Systems (4 units)

Introduction to the management of structured data beginning with an introduction to database models including relational, hierarchical, and network approaches.

Probability and Statistics Using Python (4 units)

Distribution over the real line; independence, expectation, variance, correlation. Central limit theorem. Chernoff/Hoeffding bound. Statistical tests. The Bonferroni correction.

Spring Quarter-Year One

Machine Learning (4 units)

Broad introduction to the practical side of machine-learning and data analysis. Topics in supervised learning: k-nearest neighbor classifiers, decision trees, boosting and perceptrons; and topics in unsupervised learning: k-means, PCA and Gaussian mixture models.

Data Analysis Using Hadoop and Spark (4 units)

Map-reduce, streaming analysis, and external memory algorithms and their implementation using the Hadoop and its eco-system: HBase, Hive, Pig and Spark. Projects include analyzing large existing databases.

Fall Quarter-Year Two

Data Integration & ETL (4 units)

Fundamentals of data integration and includes: schema mapping and matching, entity disambiguation, ontology development and management, data provenance, and crowd sourcing and machine learning as strategies for integration. Hands-on projects.

Elective (4 units)

Please see our website for our list of possible electives.

Winter Quarter-Year Two

Data Science Design Capstone Project (2 units)

A team design project in the final two quarters of the program culminates in a final report and an oral presentation of the capstone project. Demonstration of the working prototype.

Elective (4 units)

Please see our website for our list of possible electives.

Spring Quarter-Year Two

Data Science Design Capstone Project – Continued (2 units)

Team design project comes to completion and presented.