

INTEGRATED DATA SCIENCE

Degree Type: Certificate

The past decade has witnessed an accelerating growth in the volume and complexity of data in many data-enabled science and engineering (DESE) fields. To maximize the discovery potential, we must employ advanced data analytics methods and algorithms, visualization techniques, and high-performance computing. We are faced with unprecedented and multi-faceted challenges making skills in advanced data analytics most critical: statistics, data mining, machine learning, signal/image processing and visualization, data management and programming are becoming essential for many areas of science and engineering research. These skills bridge several disciplines and push research frontiers: from the methods disciplines of computer science, electrical engineering, applied math, and statistics to domain disciplines across science and engineering. Furthermore, the non-academic professional sector already has high demands for data scientists and engineers with PhD/MS education. They are being recruited as high-level staff at national research labs and centers, as “data wizards” at non-profit organizations, from within the financial sector, and at many industries including online retailers, social media, healthcare, and pharmaceuticals. These professional sectors are looking for data-analytics experts who can not only answer questions quantitatively, but pose the questions no one has yet identified; the same technical skills needed for data-enabled science and engineering research are in great demand in the growing number of data intensive industries as well.

The certificate in Integrated Data Science aims to organize and recognize students in this area, integrating curriculum in methods and domain disciplines and offering students the wide range of education they need to succeed as data scientists inside and outside of academia. The graduate certificate curriculum aligns well with the Northwestern Data Science Initiative and allows for further expansion, as other units across the university develop and add courses to the curriculum.

Additional resources:

- Certificate website (<https://www.ideas.ciera.northwestern.edu/certificate/>)

Certificate Offered

- Integrated Data Science Certificate (<https://catalogs.northwestern.edu/tgs/integrated-data-science/integrated-data-science-cert/>)

Integrated Data Science Courses

DATA_SCI 401-0 Data-Driven Research in Physics, Geophysics, and Astronomy (1 Unit)

Major projects in earth sciences, physics, and astronomy have revolutionized research in these fields and have created major data challenges. In this course we will review the science motivation and goals and the relevant data challenges of the Earthscope, aLIGO, and LSST projects that represent large-scale investments in these research communities. Although the goals for the three projects may appear to overlap only partially, there are strong intellectual bridges and shared challenges because of the data-intensive science involved.

DATA_SCI 421-0 Integrated Data Analytics I (1 Unit)

Data analysis in the modern age requires familiarity of many concepts and methods from statistics. This course provides an introduction to the

basics as well as exposure to some of the most advanced techniques. The emphasis will be on practical problems from physics and astronomy, rather than on theory or on statistical methods from other fields. Prior knowledge of statistics is not required.

DATA_SCI 422-0 Mathematical Inverse Methods in Earth and Environmental Sciences (1 Unit)

Theory and application of inverse methods to gravity, magnetotelluric, seismic waveform, multilateration, and students' data. Nonlinear, linearized; underdetermined, and mixed-determined problems and solution methods, such as regularized least-squares and neighborhood algorithms.

DATA_SCI 423-0 Machine Learning: Foundations, Applications, and Algorithms (1 Unit)

From robotics, speech recognition, and analytics to finance and social network analysis, machine learning has become one of the most useful set of scientific tools of our age. With this course we want to bring interested students and researchers from a wide array of disciplines up to speed on the power and wide applicability of machine learning. The ultimate aim of the course is to equip you with all the modeling and optimization tools you'll need in order to formulate and solve problems of interest in a machine learning framework. We hope to help build these skills through lectures and reading materials which introduce machine learning in the context of its many applications, as well as by describing in a detailed but user-friendly manner the modern techniques from nonlinear optimization used to solve them.