EN.685 (DATA SCIENCE)

Courses

EN.685.621. Algorithms for Data Science. 3 Credits.

This course provides a survey of computer algorithms, examines fundamental techniques in algorithm design and analysis, and develops problem-solving skills required in all programs of study involving data science. Topics include advanced data structures for data science (tree structures, disjoint set data structures), algorithm analysis and computational complexity (recurrence relations, big-O notation, introduction to complexity classes (P, NP and NP-completeness)), data transformations (FFTs, principal component analysis), design paradigms (divide and conquer, greedy heuristic, dynamic programming), and graph algorithms (depth-first and breadth-first search, ordered and unordered trees). Advanced topics are selected from among the following: approximation algorithms, computational geometry, data preprocessing methods, data analysis, linear programming, multi-threaded algorithms, matrix operations, and statistical learning methods. The course will draw on applications from Data Science. Course Prerequisite(s): EN.605.201 Introduction to Programming Using Java or equivalent. EN.605.203 Discrete Mathematics or equivalent is recommended. Course Note(s): This required foundation course must be taken before other 605.xxx courses in the degree. This course does not satisfy the foundation course requirement for Bioinformatics, Computer Science, or Cybersecurity. Students can only earn credit for one of EN.605.620, EN.605.621, or EN.685.621.

EN.685.648. Data Science. 3 Credits.

This course will cover the core concepts and skills in the interdisciplinary field of data science. These include problem identification and communication, probability, statistical inference, visualization, extract/transform/load (ETL), exploratory data analysis (EDA), linear and logistic regression, model evaluation and various machine learning algorithms such as random forests, k-means clustering, and association rules. The course recognizes that although data science uses machine learning techniques, it is not synonymous with machine learning. The course emphasizes an understanding of both data (through the use of systems theory, probability, and simulation) and algorithms (through the use of synthetic and real data sets). The guiding principles throughout are communication and reproducibility. The course is geared towards giving students direct experience in solving the programming and analytical challenges associated with data science. The assignments weight conceptual (assessments) and practical (labs, problem sets) understanding equally. Prerequisite(s): A working knowledge of Python scripting and SQL is assumed as all assignments are completed in Python.

EN.685.652. Data Engineering Principles and Practice. 3 Credits.

Data Engineering is the ingestion, transformation, storage and serving of data in ways that enable data scientists or applications to use and derive insights from data. In this course, we will look at various file-based data formats, data collection, data cleansing, data transformation, and data modeling for both relational and NoSQL databases. The course will also cover movement of data into data warehouses and/or data lakes using pipelines and workflow automation. Finally, we will discuss data security, governance, and compliance. The format of this course will be a mix of lectures, hands-on demos, and labs. Upon completing this course, students will have a deeper understanding of what a data engineer does and the various technologies that make up data engineering, along with hands-on experience working with various tools and processes.

EN.685.795. Capstone Project in Data Science. 3 Credits.

This course permits graduate students in data science to work with a faculty mentor to explore a topic in depth or conduct research in selected areas. Requirements for completion include submission of a significant paper or project. Prerequisite(s): Seven data science graduate courses including two courses numbered 605.7xx or 625.7xx or admission to the post-master's certificate program. Students must also have permission of a faculty mentor, the student's academic advisor, and the program chair.

EN.685.801. Independent Study in Data Science I. 3 Credits.

This course permits graduate students in data science to work with a faculty mentor to explore a topic in depth or conduct research in selected areas. Requirements for completion include submission of a significant paper suitable to be submitted for publication. Prerequisite(s): Seven data science graduate courses including two courses numbered 605.7xx or 625.7xx or admission to the post-master's certificate program. Students must also have permission of a faculty mentor, the student's academic advisor, and the program chair.