EN.685 (DATA SCIENCE)

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. 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.
This course will cover the core concepts and skills for data engineering with a focus on practical use cases. Data Engineering focuses on the ingestion, storage, transformation, and access of data in ways that enable data science applications to use and derive insight from data. Some of the topics that this course will touch on are dimensional modeling of data, non-relational data, data lakes, modern data warehouses, as well as different data modalities. The course will also cover some of the core supporting technologies in the data engineering world: data pipelines, containerization, schedulers, cloud technologies, and modern data engineering tools/stacks. The course is geared towards giving students direct experience in building solutions to problems associated with data engineering. The assignments will focus on the practical application of principles (labs, projects, assignments) while underscoring the understanding of the fundamental principles (assessments). Prerequisite(s): A working knowledge of Python scripting.