Data Science Graduate Program Curriculum Map

SEMESTER 1

- DATS 6101 Intro to Data Science
- DATS 6102 Data Warehousing
- DATS 6103 Intro to Data Mining

SEMESTER 2

- DATS 6202 Machine Learning I
- DATS 6401 Data Visualization
- Elective 1

SEMESTER 3

- Elective 2
- Elective 3
- Elective 4

SEMESTER 4

- Capstone

Pathway to Becoming a Data Scientist

Machine Learning Modeling

Choose from the following electives:

- DATS 6303 Deep Learning
- DATS 6312 Natural Language Processing
- DATS 6313 Time Series Analysis & Modeling
- DATS 6001 Algorithm Design for Data Science

Pathway to Becoming a Data Engineer

Machine Learning Operations (MLOps)

Choose from the following electives:

- DATS 6450 Cloud Computing
- DATS 6311 Bayesian Methods
- DATS 6450 Network Data Science

* With permission of the Data Science Director, students can also take courses across the University in areas such Geospatial Analysis (GIS), Applied Economics, Statistics, Biostatistics, Computer Science and more.

DATS 6001: Algorithm Design
This course covers Algorithm Design. Unlike the ones offered in most CS departments, this course is particularly tailored for non-CS major students. Specifically, we will only focus on (the theory and implementation of) the most important problems in algorithm design. The main goal of this course is to teach students to write code that is bug-free and has the lowest time complexity (i.e., uses the minimum time) and space complexity (uses the minimum space). In this course we will cover Data Structures (Array, Stack, Queue, and Tree) and Algorithms (Search, Sort, and Dynamic Programming).
Prerequisites: This course will use Python exclusively. It is assumed that students have used Python previously hence we will not discuss the syntax of the language in class.

DATS 6101: Intro to Data Science
Called the "Sexiest Job Title of the 21st Century" by the Harvard Business Review, Data Science and analytics are a booming industry. But what is a Data Scientist, what do they do, and how do you become one? These questions and more will be discussed and answered in this introductory course. This class covers the basic ideas and techniques of data science, including its definition and the context in data-driven computation and practical applications.
Prerequisites: None.

DATS 6102: Data Warehousing
The emergence of big data storage needs has driven adoption and development of a new class of non-relational databases commonly referred to as NoSQL databases. This course will explore the origins and the characteristics of NoSQL that distinguishes them from the traditional relational database management systems (RDMS). We will take a closer look at one database from each of the following NoSQL data models (Key-value, document and graph). We will also cover Hadoop infrastructure and HDFS storage system.
Prerequisites: None.

DATS 6103: Intro to Data Mining
This course is an introductory course on data mining. It introduces the basic concepts, principles, methods, implementation techniques, and applications of data mining, with a focus on Python and data mining algorithms for the Data Science Program. The objective of the course is to give students an overview of data mining techniques and skills to explore, analyze, and leverage data. Due to the diversity of subjects that comprise this emerging field, the class will necessarily have more breadth than depth. At the beginning of the course we will cover python to perform pre-processing and data wrangling then in the next half of the course we will cover 'core' data mining topics, such as regression and classification techniques. Students will use Python to complete the homework, assignments and projects through the course.
Prerequisites: None.

DATS 6202: Machine Learning I
In this course we will discuss the idea, practice and math of popular Machine Learning methods. While we will dive deep into the math behind some shallow and deep models, the real focus of this course is to teach students how to use popular Machine Learning tools to solve real-word problems. In this course we will use Jupyter Notebook for coding and Google Collaboratory for running the code.
Prerequisites: DATS 6101 and DATS 6103.

DATS 6303: Deep Learning

The main focus of this course will be the implementation of deep learning techniques on GPUs. Three key deep learning architectures will be covered. Multilayer Perceptrons, Convolution Networks and Long Short Term Memory are the main three deep network architecture. Some time will be spent on the background of each network, but the primary focus will be on implementation. In addition to discussing the three network architectures, the course will concentrate on three of the most popular deep learning frameworks: Tensorflow Basic, Tensorflow Advance and Pytorch. The strategy will be to present a deep network architecture, and then describe how that network can be trained and analyzed within a particular framework. Each network will be trained in a different framework.

Prerequisites: DATS 6101. Recommended Background: Prior completion of any one of MATH 2233 or equivalent; time series modeling and analysis; machine learning; or linear algebra and stochastic system.

DATS 6311: Bayesian Methods in Data Science

Introduction to Bayesian data analysis. Parameter estimation (using formal analysis, grid approximation, and Markov chain Monte Carlo), hierarchical models, generalized linear models, JAGS, and Stan. **Prerequisites:** DATS 6101 and DATS 6103.

DATS 6312: Natural Language Processing

This course is an introduction to Natural Language Processing and its basic techniques and methods. The objective of the course is to provide students an overview of natural language processing techniques that can be used to explore, analyze, and leverage natural language data stored in text. This course covers commonly used text analysis techniques and tools. Students will use Python and various packages to complete the projects through the course.

Prerequisites: DATS 6202.

DATS 6313 Time Series Analysis and Modeling

The main focus of this course is to understand, analyze, model & predict time-series dataset. In this course fundamental concepts of stochastic systems, estimation theory, time series analysis and model validation will be discussed. The course has several important keys as follows: Random variables, random processes and density function, conditional density, biased & unbiased estimators, time series analysis, model validation, auto-correlation function, partial auto-correlation and generalized partial auto-correlation function and implementation nonlinear optimization of algorithm. Python program will be mainly used throughout the course.

Prerequisites: DATS 6101 or permission of the instructor.

DATS 6401: Data Visualization

Today vast amounts of raw and refined information can be supplied and accessed to support analysis and decisions. Indeed, information access and retrieval are considered less a problem and at times a burden. The most pressing need now is to be able to present the information in a manner that is usable. This requires that cogent information be provided in context. Tothe maximum extent possible the information must be display in an intuitive manner that supports not only analytical but cognitive processes. (Source: Taylor Connor Associates LLC, all rights reserved). Tosupport this burgeoning technology requirement designers and developers of information systems need to stay current not only with the technology but the unique aspects of information visualization design. This course is intended to develop an awareness of the design concepts; an understanding of the underlying technology; introduce students to some of the currently available technologies; guide students in design protocols; and examine typical applications of those technologies. **Prerequisites:** DATS 6101, 6102 and DATS 6103.

DATS 6450 Special Topic: Cloud Computing
Analyze the different cloud computing service and deployment model and the capabilities provided by major cloud providers. You will learn how to evaluate key cloud computing services, apply cloud computing services in data science projects, and analyze the different types of cloud computing services, deployment models, and cloud vendors.
Prerequisites: DATS 6101 or permission of the instructor

DATS 6450 Special Topic: Network Data Science
Analyze critically the use of networks in data science, synthesize new concepts in network science, apply network concepts to real-world datasets, evaluate the value added by network science thinking in Data Science.
Prerequisites: DATS 6101 or permission of the instructor.

DATS 6450 Special Topics: Computer Science Foundations
This course teaches the fundamentals of computer programming utilizing Python language and presents the foundations of Computer Science. This course will cover basic computer architecture and software components. During the course of this class, students will design, write and debug Python programs. The class will use Python 3 programming language.
Prerequisites: DATS 6101 or permission of the instructor.

DATS 6450 Special Topics: Data Science Ethics
This course introduces a range of ethical topics that are relevant to today's data scientists, starting with the development of Fair Information Practice and the regulation of human subject experimentation, through industrial accidents and disasters, and the unintended consequences of the information revolution. In class, students will discuss famous cases involving ethics, crime and intellectual property. In the lab, students will create and analyze machine learning systems that clearly exhibit bias, even when bias was not part of the programming. In the second half of the course, students explore the range of controls that have been proposed to address the need to incorporate ethics into the design process.
Prerequisites: DATS 6101 or permission of the instructor.

DATS 6501: Capstone

The goal of the Capstone Project is for the students to apply the knowledge acquired during the Data Science program to a project involving actual real-world problems and data in a realistic setting. During the project, students engage in the entire process of solving a real-world data science project, from establishing a problem statement and project plan to collecting and processing actual data to applying suitable and appropriate analytic methods to the problem. Course Instructor will need to approve the problem statement and project plan before students proceed to the data collection phase. The goal of the Capstone Project is to apply theoretical knowledge gained during the time at Data Science program into a realistic project that involves a real datasets. During the project, students are heavily involved in the process of finding areal-world data science problems and solving it. The Capstone Project begins from collecting data and processing it in order to implement the appropriate analytic methods that they learned in the program to the real-world problems. In this process, problem statements and definitions are playing major role in the Capstone and the datasets can be collected from industry, government, non-governmental organizations (NGOs), or academic research. Students will work individually on a problem statement, typically specified by a faculty or the sponsor. The sponsor will usually be responsible for supplying the relevant data set. Research groups at GWU may propose projects. A list of possible projects will be posted in the blackboard so students can familiarize themselves with problems and find their interests. By approval of Course Instructor, students are free to find their own problem statement and use their own data set. The final problem statements and the datasets will need a approval by the Course Director.

Prerequisites: Students may enroll in the Capstone course upon completion of all GWU Curriculum courses or during the same semester a student is completing the last of these requirements.