

Online Master's in Data Science Curriculum

The Online Master's in Data Science program offers a thoughtfully integrated curriculum that provides students with the technical skills, quantitative expertise, and analytical insights in high demand by industry professionals. Consisting of 13 courses and totaling 30 credit hours, the program is structured so that students complete six credits per semester over five consecutive semesters. This design ensures a comprehensive and in-depth educational experience.

FALL
SEMESTER 1

- R FOR DATA SCIENCE
Credits: 1.5
- PYTHON FOR DATA SCIENCE
Credits: 1.5
- PROBABILITY & STATISTICS
Credits: 3

SPRING
SEMESTER 2

- INTRODUCTION TO MACHINE LEARNING
Credits: 3
- LINEAR MODELS
Credits: 3

SUMMER
SEMESTER 3

- STORYTELLING & COMMUNICATIONS FOR DATA SCIENCE
Credits: 1.5
- ETHICS & POLICY
Credits: 1.5
- DATABASES & DATA ARCHITECTURES
Credits: 3

FALL
SEMESTER 4

- ADVANCED MACHINE LEARNING
Credits: 3
- BEHAVIORAL DATA SCIENCE
Credits: 2
- ADVANCED DATA VISUALIZATION
Credits: 1

SPRING
SEMESTER 5

- DATA SCIENCE CAPSTONE
Credits: 3
- ADVANCED LINEAR MODELS
Credits: 3

Courses

Probability & Statistics for Data Science

This first-semester course builds the statistical foundations for applying data science in real-world scenarios. This course develops statistical thinking in the data science process applied to data collection, modeling, and inference. Students explore the connections between probability and modern statistical techniques such as exploratory data analysis (EDA) and A/B testing. (3 credits)

R Programming for Data Science

This course outfits students with the technical and practical skills required for working with modern data systems and technologies. Students learn to use the R programming language for data manipulation, data cleaning, regular expressions, feature engineering, visualization, and exploratory data analysis. Students will apply many of the most popular R packages for data science, such as Tidyverse, Dplyr, GGplot2, Shiny, and more. In addition, students will develop algorithmic thinking skills and apply them to real-world examples they will build on throughout the program. (1.5 credits)

Python Programming for Data Science

This course introduces students to the Python programming language and its application in data science. Students will use popular platforms such as Jupyter Notebooks to learn the practical aspects of data manipulation, data cleaning, feature engineering, visualization, and exploratory data analysis. Students will master the most foundational data science python libraries, including NumPy, Pandas, Seaborn, and Matplotlib. Students will build on the skills developed in this course throughout the program. (1.5 credits)

Introduction to Machine Learning

Building on the quantitative foundations established in the first semester, this course introduces students to the entire data science process, including data acquisition, exploratory data analysis, relevant machine learning methods, communicating results, and the ethical considerations in machine learning. Students use methods such as one-hot encoding and principal components analysis to preprocess the data. Students build, train, and test a variety of machine learning models such as logistic regression, random forests, and neural networks. Throughout the course, students implement and experiment with the concepts and methods of the data science process, and apply them to real-world datasets. (3 credits)

Linear Models

This course trains students in applied linear regression and time series modeling. Beginning with an introduction to fundamental concepts in regression model building and inference. This course delves into advanced techniques such as ridge regression and lasso and also explores time series methods such as autocorrelation, exponential smoothing, and ARIMA models. Bootstrapping and simulation techniques are used throughout this course to explore linear models. (3 credits)

Databases & Data Architectures

This course equips students with practical techniques for storing, retrieving, and processing structured, semi-structured, and unstructured data. Students will build on their existing knowledge of Python and learn how to query data with SQL from diverse sources, including relational databases, data lakes, and graph databases. They will use Apache Spark and Databricks to explore and model datasets too large to fit in memory on a single computer. In addition to data manipulation and analysis using SQL, the course also covers the operationalization of machine learning (ML) models using tools such as MLOps. Students will learn to deploy and manage ML models in production environments and monitor them for data drift. By the end of the course, students will have a deep understanding of big data technologies and be equipped to tackle complex data problems in a variety of contexts. This course is ideal for anyone interested in advancing their skills in data science or pursuing a career in this rapidly evolving field. (3 credits)

Storytelling & Communications for Data Scientists

This course is designed to develop communication skills for data scientists working in industry and business contexts. Students master the art of clear, effective, and engaging scientific and technical communications, with attention to the business necessity of translating complex technical subjects into actionable insights for a lay audience. Students analyze diverse audiences, adapt their ideas to diverse contexts, produce elegant and concise visual representations of data, and tell creative stories to drive their communication goals. Students learn how to design materials and deliver presentations that use data and storytelling together to inform and persuade. (1.5 credits)

Ethics & Policy in Data Science

In this course, students explore ethical frameworks, guidelines, and codes while considering how they integrate with the data science process. Existing research ethics standards provide a necessary but insufficient foundation for data science and data analytics, and so the goal of this course is to apply a critical lens to the standards in place and then learn how data scientists can further the mission of social good through ethical practices. Students will examine rapidly-changing technologies, conflicts, legal landscapes, and desires that emerge from new data practices. (1.5 credits)

Advanced Machine Learning

This course builds on the earlier training in machine learning and Python to build a variety of more advanced classifiers such as boosted ensembles, support vector machines, and Bayesian methods. Students go on to build and test deep learning convolutional neural networks for image recognition and natural language processing using technologies such as Pytorch and TensorFlow. (3 credits)

Behavioral Data Science

Behavioral Data Science trains students in data science topics specifically related to human behaviors. Major focus areas include latent variable models, recommender systems, and Natural Language Processing (NLP). Students will work with a variety of data models and techniques, like factor analysis, item response theory, topic modeling, and sentiment analysis. (2 credits)

Advanced Data Visualization

This course focuses on methods of advanced visualization of data for exploration, reporting, and monitoring tools. Students are introduced to computational tools for building interactive graphics and dashboards as well as commercial visualization software. The role of visualization in storytelling will be emphasized. (1 credit)

Data Science Capstone

This course teams groups of students with industry partners to solve real data science problems. Students learn skills of project and stakeholder management as their teams carry out all steps of the data science process: problem formulation, data acquisition, modeling, analysis, and communication of results. (3 credits)

Advanced Linear Models

This course examines extensions and generalizations of the linear regression model. Specifically, methods for fitting and evaluating logistic, multinomial, and count response models are presented using examples from a wide variety of fields. Students also delve into multilevel mixed-effects models that allow the analysis of data with repeated or clustered measurements. This course addresses techniques for missing data such as multiple imputation. (3 credits)