The College of Computer and Information Science (CCIS) and the Department of Electrical and Computer Engineering (ECE) jointly offer a new interdisciplinary Master of Science program in data science. This program is designed to give students a comprehensive framework for processing, modeling, analyzing, and reasoning about data. Students will engage in an extensive core intended to develop depth in computational modeling, data collection and integration, data storage and retrieval, data processing, modeling and analytics, and visualization. Students will also be given a variety of elective areas in CCIS, the College of Engineering (COE), and throughout the campus to explore key contextual areas or more complex technical applications. Successful program graduates will be well positioned to attain data scientist and data engineer positions in a fast-growing field or to progress into doctoral degrees in related disciplines.

The Master of Science in Data Science is comprised of eight courses; five core courses and three electives. The core courses are designed and developed by the CCIS and ECE faculty. Elective courses consist of graduate courses offered in CCIS, COE, and other partner colleges.

**Course Requirements**

The Master of Science in Data Science curriculum requires five core courses that represent the essential mathematical/statistical and technical knowledge for deep data analysis. These courses examine foundational programming concepts and languages, integration, collection, storage, retrieval, large-scale computing, mathematical concepts in statistics, linear algebra, and optimization, as well as visual and computational analysis, machine learning, and visualization. The courses are tailored toward technically or mathematically trained students.

The five core courses include:

- Two core courses in algorithms and data processing
- Two core courses in machine learning and data mining
- One core course in information visualization

Three elective courses are drawn from a selection of courses across Northeastern.

**Learning Outcomes**

Students who complete the MS degree will be able to:

- Collect data from numerous sources (databases, files, XML, JSON, CSV, and Web APIs) and integrate them into a form in which the data is fit for analysis
- Use R and Python to explore data, produce summary statistics, perform statistical analyses; use standard data mining and machine-learning models for effective analysis
- Select, plan, and implement storage, search, and retrieval components of large-scale structure and unstructured repositories
- Retrieve data for analysis, which requires knowledge of standard retrieval mechanisms such as SQL and XPath, but also retrieval of unstructured information such as text, image, and a variety of alternate formats
- Match the methodological principles and limitations of machine learning and data mining methods to specific applied problems and communicate the applicability and the advantages/disadvantages of the methods in the specific problem to nondata experts
- Carry out the full data analysis workflow, including unsupervised class discovery, supervised class comparison, and supervised class prediction; Summarize, interpret, and communicate the analysis of results
- Organize visualization of data for analysis, understanding, and communication; choose appropriate visualization method for a given data type using effective design and human perception principle
- Develop methods for modeling, analyzing, and reasoning about data arising in one or more application domains such as social science, health informatics, web and social media, climate informatics, urban informatics, geographical information systems, business analytics, bioinformatics, complex networks, public health, and game design
- Manage, process, analyze, and visualize data at scale. This outcome allows students to handle data where the conventional information technology fail.

**Placement Exams**

Each incoming masters student, regardless of his or her background, takes two placement exams administered one week prior to the beginning of the semester. The two exams cover fundamentals of computer science and programming skills and basic statistics, probability, and linear algebra. If the student does not get a B or above in a part of the placement exam, then the student must take the corresponding introductory course.

- Introduction to Programming for Data Science (DS 5010) The introductory course on fundamentals of programming and data structures covers data structures (lists, arrays, trees, hash tables, etc.), program design, programming practices, testing, debugging, maintainability, data collection techniques, and data cleaning and preprocessing. This course will have a class project where the students will use the concepts they learn to collect data from the web, clean, and preprocess and ready for analysis.
- Introduction to Linear Algebra and Probability for Data Science (DS 5020) The introductory course on basics of statistics, probability, and linear algebra covers random variables, frequency distributions, measures of central tendency, measures of dispersion, moments of a distribution, discrete and continuous probability distributions, chain rule, Bayes’ rule, correlation theory, basic sampling, matrix operations, trace of a matrix, norms, linear independence and ranks, inverse of a matrix, orthogonal matrices, range and null space of a matrix, the determinant of a matrix, positive semidefinite matrices, eigenvalues and eigenvectors.

**Program Requirements**

**Required Course Work**

A grade of B or higher is required in the following courses:

### Algorithms

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS 5800</td>
<td>Algorithms</td>
<td>4</td>
</tr>
<tr>
<td>EECE 7205</td>
<td>Fundamentals of Computer Engineering</td>
<td>4</td>
</tr>
</tbody>
</table>

### Data Management and Processing

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS 5110</td>
<td>Introduction to Data Management and Processing</td>
<td>4</td>
</tr>
<tr>
<td>Machine Learning and Data Mining</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Science, MS

DS 5220 Supervised Machine Learning and Learning Theory 4
DS 5230 Unsupervised Machine Learning and Data Mining 4

Presentation and Visualization
DS 5500

Electives
Complete 12 semester hours from the following:

**College of Computer and Information Science**
- CS 6200 Information Retrieval
- CS 5100 Foundations of Artificial Intelligence
- CS 6120 Natural Language Processing
- CS 5750 Social Computing
- CS 6350 Empirical Research Methods
- CS 7180 Special Topics in Artificial Intelligence
- CS 7280 Special Topics in Database Management

**College of Engineering**
- CIVE 7388 Special Topics in Civil Engineering
- EECE 5639 Computer Vision
- EECE 5640 High-Performance Computing
- EECE 7355 Detection and Estimation Theory
- EECE 7337 Information Theory
- EECE 7360 Combinatorial Optimization
- EECE 7370 Advanced Computer Vision
- EECE 7397 Advanced Machine Learning
- IE 5640 Data Mining for Engineering Applications
- IE 7275 Data Mining in Engineering
- IE 7280 Statistical Methods in Engineering

**College of Social Sciences and Humanities**
- PPUA 5261 Dynamic Modeling for Environmental Decision Making
- PPUA 5262 Big Data for Cities
- PPUA 5263 Geographic Information Systems for Urban and Regional Policy
- PPUA 5266 Urban Theory and Science
- PPUA 7237 Advanced Spatial Analysis of Urban Systems
- POLS 7200 Perspectives on Social Science Inquiry
- POLS 7201 Research Design
- POLS 7202 Quantitative Techniques

**D’Amore-McKim School of Business**
- BUSN 6320 Business Analytics Fundamentals
- BUSN 6324 Predictive Analytics for Managers
- BUSN 6326 Introduction to Big Data and Digital Marketing Analytics

**College of Science**
- MATH 7340 Statistics for Bioinformatics
- PHYS 5116 Complex Networks and Applications
- PHYS 7305 Statistical Physics
- PHYS 7321 Computational Physics
- PHYS 7331 Network Science Data

**Bouvé College of Health Sciences**
- NRSG 5121 Epidemiology and Population Health
- PHTH 5202 Introduction to Epidemiology
- PHTH 5210 Biostatistics in Public Health
- PHTH 5224 Social Epidemiology
- GSN 5110 Game Design and Analysis
- GSN 6350 Game Analytics

*Note:* Students that take 3-credit-hour elective courses (i.e., Bouvé, CSSH courses) will register for an accompanying data science project course in the same semester (DS 8982). In order to earn this additional credit, students will be expected to work with faculty to design an additional project in line with the curricular aims of their chosen elective and the data science core learning outcomes.

**Program Credit/GPA Requirements**
32 total semester hours required
Minimum 3.000 GPA required