Data Science and Engineering Program

Herbert G. Kayser Professor Zhigang Zhu, Program Co-Director • NAC 8/211
• Tel: 212-650-8799
Associate Professor Michael Grossberg, Program Co-Director • NAC 7/311 • Tel: 212-650-6148
Professor Akira Kawaguchi, Chair • Department Office: NAC 8/106 • Tel: 212-650-6631

General Information

This program offers Master's Degree in Data Science and Engineering (DSE).

Data Science and Data Engineering are multidisciplinary fields that apply tools and methods drawn from computer science and statistics to other knowledge domains to make predictions and decisions as well as to derive insights from both structured and unstructured data. The Data Science and Engineering (DSE) program will provide a solid foundation in the core data science and engineering skills, which will allow students to analyze, process, visualize and apply machine learning and computational statistics to problems in engineering, scientific and other disciplines. The program is administered by the Department of Computer Science, with the support of the Grove School Engineering and the Department of Mathematics. It is targeted at students with a background in science, engineering or mathematics who wish to learn data science methodology. The core data science methodology covered in the DSE program will provide students with fundamental data science and engineering computational and statistical skills. They will apply these skills to domain by combining the core knowledge with domain knowledge acquired through two or more electives taken that domain. Finally, students will complete a mandated capstone project or thesis demonstrating their mastery of the methodology.

Requirements for Admission to the DSE Program

Students are admitted to the DSE program after completing a Bachelor's degree with at least 3.0 average in Mathematics, Science or Engineering. Applicants with degrees in other fields may qualify for admission to the program depending on their experience and academic background. The general requirements are:

- Two semesters of Calculus (preferably 3 including Vector Calculus)
- Probability and Statistics (preferably 2 semesters)
- Linear Algebra
- Programming course (preferred knowledge of Python)

Applicants are encouraged to identify CCNY DSE mentors in a domain of their interest. Evidence of a potential match will be considered during admission. Include domain interest and mention any communication with a potential CCNY mentor in the personal statement. Students with baccalaureate degrees from non-English-speaking countries must submit TOEFL/IELTS Scores: the minimum is 553 (PBT), 73 (IBT) or 6. GRE submission is optional.

M.S. (DSE) Degree Requirements

<table>
<thead>
<tr>
<th>Required Courses</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six courses (3 cr. each)</td>
<td></td>
</tr>
<tr>
<td>DSE I1020: Introduction to Data Science</td>
<td></td>
</tr>
<tr>
<td>DSE I1030: Applied Statistics</td>
<td></td>
</tr>
<tr>
<td>DSE I2100: Applied Machine Learning and Data Mining</td>
<td></td>
</tr>
<tr>
<td>DSE I2410: Data Engineering: Infrastructure and Applications</td>
<td></td>
</tr>
<tr>
<td>DSE I2450: Big Data and Scalable Computation</td>
<td></td>
</tr>
<tr>
<td>DSE I2700: Visual Analytics</td>
<td></td>
</tr>
<tr>
<td>Supervised 3-credit Project or 6-credit Thesis Course (3 cr. each)</td>
<td>3-6</td>
</tr>
<tr>
<td>DSE I9800: Capstone Project (3 cr. each)</td>
<td></td>
</tr>
<tr>
<td>DSE I9900: Advanced Capstone Thesis (3 cr. each)</td>
<td>6-9</td>
</tr>
</tbody>
</table>

Two to three additional courses from domain knowledge courses listed below or any course approved by the program (3 cr. each).

Electives Courses

Additional Requirements

The students must take a minimum of 30 credits. A minimum GPA of 3.0 is required for graduation.

Currently Approved Electives

BME I530: Biomedical Signal Processing and Signal Modeling
BME I500: Biomedical Imaging
BME I6200: Organ Transport and Pharmacokinetics
ChE I550: Interfacial Phenomena
ChE I5700: Advanced Materials Engineering
ChE I8900: Nanotechnology
CE H3200/CE 5600: Engineering Hydrology
EE I6100: Advanced Algorithms
EE I6240: Computer Graphics
EE I6210: Database Systems I
EE I6220: Machine Learning and Data Mining
EE I4733: Multimedia
EE I6730: Data Reduction in Physical Sciences
EE I6716: Computer Vision
EE I6820: Web-based Geographical Information System (Web-GIS)
EE I6804: Massively Data Parallel Programming on GPUs
EE I2200: Image Processing
EE I501: Introduction to Robotics
EE I6300: Advanced Mobile Robotics
EE I6400: Computer-Aided Digital VLSI Design
EE I6530: Biologically Inspired Computation for Engineering Problems

Total credits 30

Data Science and Engineering Course Descriptions

DSE I1020: Introduction to Data Science

This course will present a survey to Data Science and introduce some of the core data science tools. While some programming experience is required for the course, the course will include a rapid introduction to Data Science programming and the stack of tools needed to process, visualize and analyze data stack with a language such as R or Python. Students will be given a high-level survey of data engineering, visual analytics, applied statistics, machine learning, and big data. The course will illustrate this bringing them through real data sets and case studies. Prereq: intro to programming CSC102/203 or equivalent, probability and statistics, calculus, linear algebra, discrete mathematics. 3 hr./wk.; 3 cr.

DSE I1030: Applied Statistics

This course will examine real data sets from a variety of domains, examine multiple models for these data sets, assess the validity of modeling assumptions, and determine the strength of conclusions that can be drawn. Topics to be covered include: 1) inferential statistics (such as hypothesis testing and estimation in parametric and nonparametric settings, conditional inference, resampling methods, cross-validation, and multiple hypothesis testing); 2) experimental design (analysis of variance) 3) Bayesian statistics (such as prior distributions, posterior and predictive inference, and Bayesian model comparison); 4) Regression and prediction (such as elements of linear and nonparametric regression, assessment of variable importance, introduction to causal inference). The course will include project-based learning and use a statistical programming language such as R or python. A strong emphasis will be placed on the critical analysis of modeling assumptions in real-world settings. Prereq: intro to programming CSC102/203 or equivalent, probability and statistics, calculus, linear algebra, discrete mathematics. 3 hr./wk.; 3 cr.

DSE I1100: Applied Machine Learning and Data Mining

Introduction to machine learning, data mining, and statistical pattern recognition. Topics include: 1) Supervised learning (parametric/nonparametric algorithms, support vector machines, kernels, neural networks, deep learning), 2) Unsupervised learning (clustering, non-parametric techniques, dimensionality reduction); 3) Best practices in machine learning (bias/variance theory, model selection and evaluation, resampling). In this class, you will learn about the most effective machine learning techniques, and gain practice implementing them and getting them to work for yourself. More importantly, you’ll learn about not only the theoretical underpinnings of learning, but also gain the practical know-how needed to quickly and powerfully apply these techniques to new problems. Prereq. DSE I1020, Intro to Data Science and DSE I1030, Applied Statistics, or equivalents. 3 hr./wk.; 3 cr.
DSE 14200: Data Engineering: Infrastructure and Applications
This course will train students in the handling of big data sources derived from various environments including traditional business activities, web-based transactions, and social media. The course will also discuss the range of data formats, application types and emerging issues in data integration. As part of this it will introduce the range of research topics and mentors participating in the Data Science and Engineering Program and offering capstone project opportunities. The course will begin with a discussion of high-end traditional database systems focusing on query processing, crash recovery, and transaction concurrency control. This will be followed by a detailed look at object-relational databases, distributed and federated databases, and cloud-based data-warehousing. NoSQL databases (e.g., Cassandra and Neo4j) and parallel data analysis tools (e.g., Hadoop, Spark) will be introduced. The main emphasis of the course is hands-on training in state-of-the-art software development environments. Project based system development work will be an essential component of the course. Prereq. DSE 11200, Intro to Data Science and DSE 11300, Applied Statistics, or equivalents. 3 hr./wk.; 3 cr.

DSE 12700: Visual Analytics
This course will give an overview of visual analytics as well as the overlapping fields of information and scientific visualization. Students will learn to programmatically process and analyze data with Python libraries widely used in statistics, engineering, science and finance. We will cover the design of effective visualizations. Students will learn to build data visualizations directly using a variety of data visualization libraries such as matplotlib, seaborn, and bokeh (Python) and interactive web-based visual analytics using JavaScript and D3. Project groups of students will each propose, design and build a visualization of a data set. The goals of the course are for students to: (1) Recognize the appropriate applications and value of visualizations; (2) Critically evaluate visualizations and suggest improvements and refinements; (3) Apply a structured design process to create effective visualizations; (4) Use programmatic tools to scrape, clean, and process data; (5) Use principles of human perception and cognition in visual analytics design; (6) Use visual analytics and statistics tools to explore data; and (7) Create web-based interactive visualizations. Prereq. DSE 11200, Intro to Data Science and DSE 11300, Applied Statistics, or equivalents. The course also requires students have programming experience such as CSC 102/103 or equivalent. 3 hr./wk.; 3 cr.

DSE 12450: Big Data and Scalable Computation
The course aims to provide a broad understanding of big data and current technologies in managing and processing them with a focus on the urban environment. With storage and networking getting significantly cheaper and faster, big data sets could easily reach the hands of data enthusiasts with just a few mouse clicks. These enthusiasts could be policy makers, government employees or managers, who would like to draw insights and (business) value from big data. Thus, it is crucial for big data to be made available to the non-expert users in such a way that they can process the data without the need of a supercomputing expert. One such approach is to build big data programming frameworks that can deal with big data in as close a paradigm as the tools deal with "small data." Such a framework should be as simple as possible, even if not as efficient as custom-designed parallel solutions. Users should expect that if their code works within these frameworks for small data, it will also work for big data. General topics of this course include: big data ecosystems, parallel and streaming programming model, MapReduce, Hadoop, Spark, Pig, and NoSQL solutions. Hands-on labs and exercises will be offered throughout to bolster the knowledge learned in each module. Prereq. DSE 11200, Intro to Data Science and DSE 11300, Applied Statistics, or equivalents. 3 hr./wk.; 3 cr.

DSE 19800: Capstone Project
A capstone project is experimental project under the direction of a faculty advisor. All students will register and submit a project report after one semester to receive a grade. Students may work together on the same data sets and challenges but must establish separate subprojects, and submit individual reports/theses. These independent study projects should involve an analysis of a data set in an application field using statistical learning/data mining techniques such as non-linear regression, supervised/unsupervised learning, dimension reduction, reinforcement learning, collaborative filtering or big-data methodology such as map-reduce/spark. Prereq. DSE 11200, Intro to Data Science and DSE 11300, Applied Statistics, DSE 11400, Machine Learning, DSE 12400, Data Engineering. 3 hr./wk.; 3 cr.

DSE 19900: Adv Capstone Thesis
Students, with approval from their mentor, may register for a second semester to complete a second independent study (advanced capstone thesis) building on the first semester work in DSE 19800 Capstone Project. Prereq. DSE 11200, Intro to Data Science and DSE 11300, Applied Statistics, DSE 11400, Machine Learning, DSE 12400, Data Engineering, and DSE 19800, Capstone Project. 3 hr./wk.; 3 cr.

Data Science and Engineering Faculty
Asheel Amarasingham, Associate Professor, Mathematics
M.S., Ph.D. (Applied Mathematics), Brown University; BSc (Mathematics, Cognitive Science), U. Virginia

Marom Bikson, Professor, Biomedical Engineering and Bioinformatics
Ph.D. (BME), Case Western Reserve University; B.S. (BME), Johns Hopkins University

Peter Brass, Professor, Computer Science
Dipl. Math, Dr. rer. nat. (Math), Technical Univ. of Braunschweig

Shirshendu Chatterjee, Assistant Professor, Mathematics
M.S., Ph.D. (Operations Research & Information Engineering), Cornell Univ.; B.Stat, M.Stat (Statistics), Indian Statistical Institute, India

Alexander Couzis, Professor, Chemical Engineering
Ph.D., M.S. (ChemEng), Univ. Michigan; B.S. (ChemEng) NTU Athens (Greece)

Ronak Etemadpour, Assistant Professor, Computer Science
M.S. (C.Sc.), Universiti Sains Malaysia (Malaysia); Ph.D. (C.Sc.), Jacobs University Bremen (Germany)

Nelly Fazio, Associate Professor, Computer Science
Laurea (C.Sc.), Universita di Catania (Italy); M.Sc. (C.Sc.), New York University, Ph.D. (C.Sc.)

Balazs Fekete, Assistant Professor, Civil Engineering
M.Sc. (Civil Eng), Budapest Univ.; Ph.D. (Earth Sciences), Univ New Hampshire

Bingmei Fu, Professor, Biomedical Engineering and Bioinformatics
B.S. (BioSci), Modern Mechanics, Univ. of Sci. & Tech. of China; Ph.D. (Mechanical Engineering), The City University of New York

Rosario Gennaro, Professor, Computer Science
Laurea (C.Sc.), Universita di Catania (Italy); M.Sc. (C.Sc.), Massachusetts Institute of Technology, Ph.D. (C.Sc.)

Lanek Gilchrist, Associate Professor, Chemical Engineering
Ph.D., UC Davis; B.S., Louisiana State Univ.

Irina Gladkova, Associate Professor, Computer Science
B.S. (Mathematics), Donets State Univ.; Ph.D. (Mathematics) CUNY

Michael D. Grossberg, Assistant Professor, Computer Science
B.A., Univ. of Penn.; Ph.D., MIT

Akira Kagawauchi, Professor and Chair, Computer Science
B.S. (Admin. Engr.), Keio Univ. (Japan); M.S.; M.Sc. (C.Sc.), Columbia Univ., Ph.D. (C.Sc.)

Nir Krakauer, Assistant Professor, Civil Engineering
Ph.D. (Geochemistry), California Institute of Technology

Ilona Kretzschmar, Professor, Chemical Engineering
Ph.D. (Gas Phase Chemistry), TU Berlin; Diploma (Chemistry), TU Berlin (Germany)

Charles Maldarelli, Professor, Chemical Engineering
DSE, M.S., B.S., Columbia University

Abbe Mowshowitz, Professor, Computer Science
B.S. (Math), Univ. of Chicago; M.S. (Math), Univ. of Michigan, Ph.D. (C.Sc.)

Lucas C. Parra, Professor, Biomedical Engineering and Bioinformatics
B.S. (Physics and CS), Ph.D. (Physics), Ludwig Maximilian University, Munich

Sang-Woo Seo, Associate Professor, Electrical Engineering
B.S., Ajou Univ. (S. Korea); M.S., Kwangju Inst. Of Sci and Tech. (S. Korea); Ph.D. (ECE), Georgia Tech.

William E. Skeith, Associate Professor, Computer Science
B.S. (Math), Pepperdine Univ., BA (C.Sc.); Univ. of Cal. Los Angeles, M.A. (Math), Ph.D.

Hansong Tang, Associate Professor, Civil Engineering
D.Sc. Peking Univ.; Ph.D. Georgia Inst. Tech

YingLi Tian, Professor, Electrical Engineering
B.S., Tian Jin Univ. (China); M.S., Ph.D. (EE), Univ. of Hong Kong

Raymond Tu, Associate Professor, Chemical Engineering
B.S., U Florida; Ph.D., UC Santa Barbara

M. Umit Uyar, Professor, Electrical Engineering
B.S., Istanbul Teknik Univ. (Turkey); M.S., Ph.D. (EE), Cornell Univ.

Huy T. Vo, Assistant Professor, Computer Science
B.S. (C.Sc.), University of Utah, Ph.D. (C.Sc.)

Jie Wei, Professor, Computer Science
B.S. (C.Sc.), Univ. of Sci. & Tech. of China (China); M.S. (C.Sc.), Chinese
George Wolberg, Professor, Computer Science
B.E. (EE), Cooper Union, M.E. (EE), Ph.D. (C.Sc.), Columbia Univ.

Jizhong Xiao, Professor, Electrical Engineering
B.S., East China Inst. of Tech. (China); M.S. (EE), Nanyang Tech. Univ. (Singapore); Ph.D. (ECE), Michigan State Univ.

Bo Yuan, Assistant Professor, Electrical Engineering
B.S., M.E., Nanjing University (China); Ph.D., University of Minnesota

Jianting Zhang, Associate Professor, Computer Science
B.S. (Water Resources and Environment), Nanjing Univ. (China); M.S. (Physical Geography), Nanjing Univ. (China); M.S. (CSc), Univ. of Oklahoma; Ph.D. (C.Sc.), Univ. of Oklahoma

Zhigang Zhu, Herbert G. Kayser Professor, Computer Science
B.S., (C.Sc.), Tsinghua Univ., M.E., Ph.D.