Data Science

Program Description

The Master of Data Science (non-thesis) program is designed to give candidates a foundation in statistics and computer science and also provide knowledge in a particular application domain of science or engineering. The balance between these three elements is a strength of the program and can prepare candidates for data science careers in industry, government, or for further study at the PhD level. Throughout is an emphasis on working in teams, creative problem solving, and professional development.

The Data Science Certificates are designed for college graduates and professionals interested in the emerging field of data science as applied within their individual fields of study or industries.

Contact

Program Director: Soutir Bandyopadhyay

datascience.mines.edu

Program Coordinator: Rachel McDonald

Master of Data Science (Non-Thesis)

The field of Data Science draws on elements of computer science, statistics and interdisciplinary applications to address the unique needs of gaining knowledge and insight through data analysis. This Masters Non-Thesis program is designed to give candidates a foundation in statistics and computer science and also provide knowledge in a particular application domain of science or engineering. The balance between these three elements is a strength of the program and can prepare candidates for Data Science careers in industry, government, or for further study at the PhD level. Moreover, the coursework will be flexible and tailored to each candidate. For example, the program will allow a candidate to increase his/her skills in data analytics while developing a focused area of application or alternatively allow a candidate with depth in an area of application to gain skills in statistics and computer science.

Program Requirements

This program will follow a 3 X 3 + 1 design: three modules and a mini-module.

Modules (each consisting of three 3-credit courses)

Data Modeling and Statistical Learning

MATH530	INTRODUCTION TO STATISTICAL METHODS (Online Only) Students who have passed MATH335 should replace MATH530 with any of the following: MATH500, MATH531, MATH532, MATH533, MATH534, MATH535, MATH536, MATH537, MATH538, or MATH551	3.0
MATH560	INTRODUCTION TO KEY STATISTICAL LEARNING METHODS I (Online Only)	3.0
MATH561	INTRODUCTION TO KEY STATISTICAL LEARNING METHODS II (Online Only)	3.0

Machine Learning, Data Processing and Algorithms, and Parallel Computation

DSCI503	ADVANCED DATA SCIENCE	3.0
DSCI570	INTRODUCTION TO MACHINE LEARNING	3.0
DSCI575	ADVANCED MACHINE LEARNING	3.0

Individualized and Domain Specific Coursework

Electives for the third module can be designed by the student but the plan needs to be approved by the program curriculum committee. Although this individualized module can draw on graduate courses from across the university, two specific examples from engineering and geophysics are given below:

Electrical Engineering

EENG510	ADVANCED DIGITAL SIGNAL PROCESSING	3.0
EENG509	SPARSE SIGNAL PROCESSING	3.0
EENG511	CONVEX OPTIMIZATION AND ITS ENGINEERING APPLICATIONS	3.0
EENG515	MATHEMATICAL METHODS FOR SIGNALS AND SYSTEMS	3.0
or EENG519	ESTIMATION THEORY AND KALMAN FILTERING	

Geophysics

GPGN533	GEOPHYSICAL DATA INTEGRATION & GEOSTATISTICS	3.0
GPGN570	APPLICATIONS OF SATELLITE REMOTE SENSING	3.0

Mini-module (comprised of three 1-credit courses)

Professional Development

SYGN502	INTRODUCTION TO RESEARCH ETHICS	1.0
SYGN5XX	LEADERSHIP AND TEAMWORK	1.0
LICM501	PROFESSIONAL ORAL COMMUNICATION	1.0
EDNS544	INNOV8X	3.0

Sample Course Schedule

First Year

Fall		**lec**	**lab**	**sem.hrs**
DSCI503	ADVANCED DATA SCIENCE			3.0
DSCI570	INTRODUCTION TO MACHINE LEARNING			3.0
MATH530	INTRODUCTION TO STATISTICAL METHODS			3.0
ELECT	Elective Approved by Program*			3.0
				12.0

Spring		**lec**	**lab**	**sem.hrs**
DSCI575	ADVANCED MACHINE LEARNING			3.0
MATH560	INTRODUCTION TO KEY STATISTICAL LEARNING METHODS I			3.0
ELECT	Elective Approved by Program*			3.0
LICM501	PROFESSIONAL ORAL COMMUNICATION			1.0
				10.0

Second Year				
Fall		**lec**	**lab**	**sem.hrs**
MATH561	INTRODUCTION TO KEY STATISTICAL LEARNING METHODS II			3.0
ELECT	Elective Approved by Program*			3.0
SYGN502	INTRODUCTION TO RESEARCH ETHICS			1.0
SYGN5XX	LEADERSHIP AND TEAMWORK			1.0
				8.0

Total Semester Hrs: 30.0

*Electives for the third module can be designed by the student but the plan needs to be approved by the Data Science program curriculum committee. This individualized module can draw on graduate courses from across the university.

Mines Combined Undergraduate / Graduate Degree Program

Students enrolled in Mines' combined undergraduate/graduate program may double count up to six credits of graduate coursework to fulfill requirements of both their undergraduate and graduate degree programs. These courses must have been passed with "B-" or better, not be substitutes for required coursework, and meet all other University, Department, and Program requirements for graduate credit.

Students are advised to consult with their undergraduate and graduate advisors for appropriate courses to double count upon admission to the combined program.

Master of Data Science (Non-Thesis) Online

The Master of Data Science (Non-Thesis) program is designed to give candidates a foundation in statistics and computer science and also provide knowledge in a particular application domain of science or engineering. The balance between these three elements is a strength of the program and can prepare candidates for Data Science careers in industry, government, or for further study at the PhD level. Throughout is an emphasis on working in teams, creative problem solving, and professional development.

This non-thesis master's program gives students a foundation in statistics and computer science, while also providing knowledge in a particular application domain of science or engineering. It is pitched a higher level of statistics and computer science than one would encounter in a typical data analytics curriculum.

The balance between these elements is a strength of the program and prepares students for data science careers in industry, government or for further study at the PhD level. The emphasis on some foundational knowledge will prepare students to be more innovative in their approach to data analysis and not rely on simply using software packages in a standard way. Moreover, the three elective courses can be tailored to each student's interests. This program allows students to either increase their skill in data analytics while *developing a focused area of application* or alternatively to allow a student with depth in one area of application to *gain skills in statistics and computer science*.

Program

This program will follow a 2 X 3 + 1 electives module design: two core modules (18 credit hours) and a third comprised of pre-approved electives (12 credit hours), for a total of 30 credit hours.

Required Module: Data Modeling and Statistical Learning

MATH530	INTRODUCTION TO STATISTICAL METHODS Students who have passed MATH335 should replace MATH530 with any of the following: MATH500, MATH531, MATH532, MATH533, MATH534, MATH535, MATH536, MATH537, MATH538, or MATH551	3.0
MATH560	INTRODUCTION TO KEY STATISTICAL LEARNING METHODS I	3.0
MATH561	INTRODUCTION TO KEY STATISTICAL LEARNING METHODS II	3.0

REQUIRED Module: Machine Learning, Data Processing and Algorithms, and Parallel Computation

DSCI503	ADVANCED DATA SCIENCE	3.0
DSCI570	INTRODUCTION TO MACHINE LEARNING	3.0
DSCI575	ADVANCED MACHINE LEARNING	3.0

Individualized and Domain Specific Coursework

In addition to the required 18 credit hours above, students must take 12 credit hours of electives. 3 credits must pertain to professional development and the remaining are in areas of subject matter or data science methods. Electives can be designed by the student, but the plan needs to be approved by the program curriculum committee in advance of taking the courses. Although this individualized module can draw on graduate courses from across the university that are relevant to the students' career or focused field, a few specific examples to form a foci in electrical engineering and geophysics are given below. Other course options are listed below these two examples. Also, some examples of professional development courses are included in the last section. The list of courses is not comprehensive, but rather gives illustrative examples of appropriate foci and courses. Some electives may not be available in online or in-person modality, check the bulletin or your advisor for updated information.

ELECTIVES: Electrical Engineering Foci

EENG509	SPARSE SIGNAL PROCESSING	3.0
EENG510	ADVANCED DIGITAL SIGNAL PROCESSING	3.0
EENG515	MATHEMATICAL METHODS FOR SIGNALS AND SYSTEMS	3.0
EENG519	ESTIMATION THEORY AND KALMAN FILTERING	3.0

ELECTIVES: Geological & Geophysics Foci

GPGN533	GEOPHYSICAL DATA INTEGRATION & GEOSTATISTICS	3.0
GPGN570	APPLICATIONS OF SATELLITE REMOTE SENSING	3.0
GEGN575	APPLICATIONS OF GEOGRAPHIC INFORMATION SYSTEMS	3.0

GEGN579	PYTHON SCRIPTING FOR GEOGRAPHIC INFORMATION SYSTEMS	3.0
GEOL557	EARTH RESOURCE DATA SCIENCE 1: FUNDAMENTALS	3.0
GEOL558	EARTH RESOURCE DATA SCIENCE 2: APPLICATIONS AND MACHINE-LEARNING	3.0

ELECTIVES: Other Subject Area Foci

EBGN525	BUSINESS ANALYTICS	3.0
EBGN559	SUPPLY CHAIN ANALYTICS	3.0
EBGN560	DECISION ANALYTICS	3.0
EBGN571	MARKETING ANALYTICS	3.0
PEGN551	PETROLEUM DATA ANALYTICS - FUNDAMENTALS	3.0

Certificate Programs in Data Science

Program Requirements

There are five Certificates in Data Science. Applicants for each are required to have an undergraduate degree to be admitted into the Certificate programs. Course prerequisites, if any, are noted for each Certificate program.

Students working toward one of the Data Science Certificates are required to successfully complete 12 credits, as detailed below for each Certificate. The courses taken for the Certificates can be used towards a Master's or PhD degree at Mines, however courses used for one Data Science Certificate cannot also be counted toward another Data Science Certificate.

Graduate Certificate in Data Science - Statistical Learning (12 credits)

The Data Science - Statistical Learning Graduate Certificate is an online or residential program focusing on statistical methods for interpreting complex data sets and quantifying the uncertainty in a data analysis. The Certificate also includes gaining new skills in computer science but is grounded in statistical models for data, also termed statistical learning, rather than algorithmic approaches. Students will develop an essential skill set in statistical methods most commonly used in data science along with the understanding of the methods' strengths and weaknesses. Moreover, the coursework will cover a broad range of applications making it relevant for varied scientific and engineering domains.

Applicants must have completed the following courses, or their equivalents, with a B- or better: CSCI261 and CSCI262 Data Structures, MATH332 Linear Algebra and MATH334 Introduction to Probability.

DSCI503	ADVANCED DATA SCIENCE	3.0
DSCI530	STATISTICAL METHODS I	3.0
DSCI560	INTRODUCTION TO KEY STATISTICAL LEARNING METHODS I	3.0
DSCI561	INTRODUCTION TO KEY STATISTICAL LEARNING METHODS II	3.0

Graduate Certificate in Data Science - Earth Resources (12 credits)

The Graduate Certificate in Data Science - Earth Resources is an online program building on the foundational concepts in data science as it pertains to managing surface and subsurface Earth resources and on specific applications (use cases) from the petroleum and minerals industries as well as water resource monitoring and remote sensing of Earth change. The Certificate includes one core introductory Data Science course, two courses specific to Earth resources and one elective.

DSCI503	ADVANCED DATA SCIENCE	3.0
GEOL557	EARTH RESOURCE DATA SCIENCE 1: FUNDAMENTALS	3.0
GEOL558	EARTH RESOURCE DATA SCIENCE 2: APPLICATIONS AND MACHINE-LEARNING	3.0
ELECTIVE	(1) ELECTIVE FROM LIST BELOW	3.0

Graduate Certificate in Data Science - Earth Resources Electives (select ONE (1) from the list below):

Geospatial Focus:		
GEGN575	APPLICATIONS OF GEOGRAPHIC INFORMATION SYSTEMS	3.0
GEGN579	PYTHON SCRIPTING FOR GEOGRAPHIC INFORMATION SYSTEMS	3.0
Petroleum Focus:		
GPGN519	ADVANCED FORMATION EVALUATION	3.0
GPGN547	PHYSICS, MECHANICS, AND PETROPHYSICS OF ROCKS	3.0
GPGN558	SEISMIC DATA INTERPRETATION AND QUANTITATIVE ANALYSIS	3.0
GPGN651	ADVANCED SEISMOLOGY	3.0
PEGN522	ADVANCED WELL STIMULATION	3.0
PEGN551	PETROLEUM DATA ANALYTICS - FUNDAMENTALS	3.0
Mining Focus:		
MNGN548	INFORMATION TECHNOLOGIES FOR MINING SYSTEMS	3.0
Hydrology Focus:		
CEEN581	WATERSHED SYSTEMS MODELING	3.0
Additional Options:		
DSCI/MATH530	STATISTICAL METHODS I	3.0
EBGN525	BUSINESS ANALYTICS	3.0

Graduate Certificate in Petroleum Data Analytics (12 credits)

The Graduate Certificate in Petroleum Data Analytics is an online program building on the foundational concepts in statistics and focusing on the data foundation of the oil and gas industry, the challenges of Big Data to oilfield operations and on specific applications (use cases) for petroleum analytics. The Certificate includes two core introductory Data Science courses and two course specific to petroleum engineering.

DSCI503	ADVANCED DATA SCIENCE	3.0
DSCI530	STATISTICAL METHODS I	3.0
PEGN551	PETROLEUM DATA ANALYTICS - FUNDAMENTALS	3.0
PEGN552	PETROLEUM DATA ANALYTICS - APPLICATIONS	3.0

Graduate Certificate in Business Analytics

```
The certificate is an online or residential program. The
 requirements are to complete the core course and two elective
 courses:
```

Core Course

EBGN525	BUSINESS ANALYTICS	3.0

Elective Courses

EBGN559	SUPPLY CHAIN ANALYTICS	3.0
EBGN560	DECISION ANALYTICS	3.0
EBGN571	MARKETING ANALYTICS	3.0

Course substitutions may be approved on a case-by-case basis by the certificate director. Completing the certificate will also position students to apply to either the master of science in engineering and technology management degree or the master of science in data science degree, as the certificate courses can be applied to either degree.

Graduate Certificate in Data Science - Foundations (12 credits)

The Data Science - Foundations Graduate Certificate is an online or residential program focusing on the foundational concepts in statistics and computer science that support the explosion of new methods for interpreting data in its many forms. The Certificate balances an introduction to data science with teaching basic skills in applying methods in statistics and machine learning to analyze data. Students will gain a perspective on the kinds of problems that can be solved by data intensive methods and will also acquire new analysis skills outside of the certificate. Moreover, the coursework will cover a broad range of applications, making it relevant for varied scientific and engineering domains.

Applicants must have completed the following courses, or their equivalents, with a B- or better: CSCI261 and CSCI262 Data Structures, MATH332 Linear Algebra and MATH334 Introduction to Probability.

DSCI503	ADVANCED DATA SCIENCE	3.0
DSCI570	INTRODUCTION TO MACHINE LEARNING	3.0
MATH530	INTRODUCTION TO STATISTICAL METHODS	3.0
MATH560	INTRODUCTION TO KEY STATISTICAL LEARNING METHODS I	

Graduate Certificate in Data Science - Computer Science (12 credits)

The Data Science - Computer Science Graduate Certificate is an online or residential program focusing on data science concepts within computer science (e.g., computational techniques and machine learning) plus prerequisite knowledge (e.g., probability and regression). The aim of this certificate is to help students develop an essential skill set in data analytics, including (1) deriving predictive insights by applying advanced statistics, modeling, and programming skills, (2) acquiring in-depth knowledge of machine learning and computational techniques, and (3) unearthing important questions and intelligence for a range of industries, from product design to finance.

Applicants must have completed the following courses, or their equivalents, with a B- or better: CSCI261 and CSCI262 Data Structures, MATH213 Calculus III and MATH332 Linear Algebra. DSCI530 Statistical Methods I, will serve as the MATH201 Probability and Statistics prerequisite for the two machine learning courses of the certificate (DSCI570 Introduction to Machine Learning and DSCI575 Machine Learning).

DSCI503	ADVANCED DATA SCIENCE	3.0
MATH530	INTRODUCTION TO STATISTICAL METHODS	3.0
DSCI570	INTRODUCTION TO MACHINE LEARNING	3.0
DSCI575	ADVANCED MACHINE LEARNING	3.0

Courses

DSCI503. ADVANCED DATA SCIENCE. 3.0 Semester Hrs.
(I, II) This course will teach students the core skills needed for gathering, cleaning, organizing, analyzing, interpreting, and visualizing data. Students will use the python programming language and related toolkits for data manipulation and the use and application of statistical and machine learning for data analysis. The course will be primarily focused on applications, with an emphasis on working with real (non-synthetic) datasets. Students will propose and design a semester project using a dataset from their domain of interest, leveraging the concepts and skills acquired from this course (e.g., data analysis, ethical considerations, evaluation and synthesis of results, storytelling and visualization). Prerequisite: CSCI200 with a grade of C- or higher or CSCI262 with a grade of C- or higher, MATH201 or MATH334 OR Graduate level standing and at least CSCI128 or equivalent.
Course Learning Outcomes

- Acquire, clean, and organize structured and unstructured data from a variety of sources, including raw data files, online repositories, and through the use of web scraping and various APIs.
- Utilize toolkits and exploration to preprocess small, medium, and large datasets for input to statistical and machine learning algorithms, including methods of feature extraction, outlier removal, and dimensionality reduction.
- Apply statistical and machine learning toolkits to small, medium, and large datasets, including applications of regression, classification, clustering, and a brief introduction to neural networks.
- Conduct analysis of results and evaluate the predictive power of various statistical and machine learning techniques.
- Develop storytelling and visualization skills to inform (exploratory) or persuade (explanatory) a specific audience using data.
- Recognize and address the ethical issues arising from data collection and statistical and machine learning.
- Design, propose, and present a semester project using a dataset from their domain of interest leveraging the concepts and skills from this course.

DSCI530. STATISTICAL METHODS I. 3.0 Semester Hrs.
Introduction to probability, random variables, and discrete and continuous probability models. Elementary simulation, data summarization and analysis using the R Data Analysis Environment. Confidence intervals and hypothesis testing for means and variances. Chi square tests. Distribution-free techniques and regression analysis. Students are expected to have knowledge of probability covered in MATH334 or an equivalent course. Prerequisite: MATH334 or equivalent.

DSCI560. INTRODUCTION TO KEY STATISTICAL LEARNING METHODS I. 3.0 Semester Hrs.
Part one of a two-course series introducing statistical learning methods with a focus on conceptual understanding and practical applications. Methods covered will include Introduction to Statistical Learning, Linear

Regression, Classification, Resampling Methods, Basis Expansions, Regularization, Model Assessment and Selection. Prerequisite: DSCI530 or MATH530.

DSCI561. INTRODUCTION TO KEY STATISTICAL LEARNING METHODS II. 3.0 Semester Hrs.
Equivalent with MATH561,
Part two of a two course series introducing statistical learning methods with a focus on conceptual understanding and practical applications. Methods covered will include Non-linear Models, Tree-based Methods, Support Vector Machines, Neural Networks, Unsupervised Learning. Prerequisite: DSCI560 or MATH560.

DSCI570. INTRODUCTION TO MACHINE LEARNING. 3.0 Semester Hrs.
(I, II) The goal of machine learning is to build computer systems that improve automatically with experience, which has been successfully applied to a variety of application areas, including, for example, gene discovery, financial forecasting, and credit card fraud detection. This introductory course will study both the theoretical properties of machine learning algorithms and their practical applications. Students will have an opportunity to experiment with machine learning techniques and apply them to a selected problem in the context of term projects. Graduate students must complete a more challenging project that utilizes complex machine learning algorithms, requiring a deeper understanding of machine learning approaches and critical thinking. Prerequisite: DSCI503.
Course Learning Outcomes

- Apply supervised, unsupervised, reinforcement machine learning models and deep learning models to solve problems in areas such as prediction, recognition and classification.
- Explore and develop with various tools, techniques and libraries in Python for data processing, feature extraction, visualization, validation and evaluation.
- Create data visualization tools, techniques, and libraries in Python to visualize high dimensional or complex data for stakeholders.
- Determine ethical implications through interpretability of big data and results from the application of various machine learning models.
- Design and develop a machine learning product that solves their chosen real-world challenge.
- Create a video presentation that succinctly outlines the problem, solutions, conclusions, and lessons learned regarding product development for the stakeholders.

DSCI575. ADVANCED MACHINE LEARNING. 3.0 Semester Hrs.
The goal of machine learning research is to build computer systems that learn from experience and that adapt to their environments. Machine learning systems do not have to be programmed by humans to solve a problem; instead, they essentially program themselves based on examples of how they should behave, or based on trial and error experience trying to solve the problem. This course will focus on the methods that have proven valuable and successful in practical applications. The course will also contrast the various methods, with the aim of explaining the situations in which each is most appropriate. Prerequisite: DSCI570.

DSCI598. SPECIAL TOPICS. 0-6 Semester Hr.

DSCI598. SPECIAL TOPICS. 1-6 Semester Hr.

DSCI598. SPECIAL TOPICS. 1-6 Semester Hr.

DSCI598. SPECIAL TOPICSE. 1-6 Semester Hr.

DSCI598. DATA SCIENCE. 1-6 Semester Hr.

DSCI599. INDEPENDENT STUDY. 0.5-6 Semester Hr.

DSCI599. INDEPENDENT STUDY. 0.5-6 Semester Hr.

Professor

Douglas Nychka, Applied Mathematics & Statistics

Associate Professors

Dorit Hammerling, Applied Mathematics & Statistics

Soutir Bandyopadhyay, Applied Mathematics and Statistics

Teaching Professor

Wendy Fisher, Computer Science

Teaching Assistant Professors

Nathan Lenssen, Applied Mathematics and Statistics

Zibo Wang, Computer Science

Research Associate Professor

Zane Jobe, Geology and Geological Engineering