Data Science

Degrees Offered
- Master of Science in Data Science (non-thesis)
- Graduate Certificate in Data Science – Statistical Learning
- Graduate Certificate in Data Science – Earth Resources
- Graduate Certificate in Petroleum Data Analytics
- Graduate Certificate in Business Analytics
- Post-Baccalaureate Certificate in Data Science – Foundations
- Post Baccalaureate Certificate in Data Science – Computer 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.

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
- 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

Machine Learning, Data Processing and Algorithms, and Parallel Computation
- DSCI503: ADVANCED DATA SCIENCE 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
- INNO544: INNOV8X CREATE DSCI 1.0

Sample Course Schedule

First Year

<table>
<thead>
<tr>
<th>Fall</th>
<th>lec</th>
<th>lab</th>
<th>sem.hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI503: ADVANCED DATA SCIENCE</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSCI570: INTRODUCTION TO MACHINE LEARNING</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSCI530: STATISTICAL METHODS I</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELECT</td>
<td>Elective Approved by Program*</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12.0</td>
</tr>
</tbody>
</table>

Spring

<table>
<thead>
<tr>
<th>lec</th>
<th>lab</th>
<th>sem.hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI575: MACHINE LEARNING</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>DSCI560: INTRODUCTION TO KEY STATISTICAL LEARNING METHODS I</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>ELECT</td>
<td>Elective Approved by Program*</td>
<td>3.0</td>
</tr>
<tr>
<td>LICM501: PROFESSIONAL ORAL COMMUNICATION</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Second Year

<table>
<thead>
<tr>
<th>lec</th>
<th>lab</th>
<th>sem.hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI561: INTRODUCTION TO KEY STATISTICAL LEARNING METHODS II</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>ELECT</td>
<td>Elective Approved by Program*</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
*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**

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI530</td>
<td>Statistical Methods I</td>
<td>3.0</td>
</tr>
<tr>
<td>DSCI560</td>
<td>Introduction to Key Statistical Learning Methods I</td>
<td>3.0</td>
</tr>
<tr>
<td>DSCI561</td>
<td>Introduction to Key Statistical Learning Methods II</td>
<td>3.0</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI503</td>
<td>Advanced Data Science</td>
<td>3.0</td>
</tr>
<tr>
<td>DSCI570</td>
<td>Introduction to Machine Learning</td>
<td>3.0</td>
</tr>
<tr>
<td>DSCI575</td>
<td>Machine Learning</td>
<td>3.0</td>
</tr>
</tbody>
</table>

**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**

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>EENG509</td>
<td>Sparse Signal Processing</td>
<td>3.0</td>
</tr>
<tr>
<td>EENG510</td>
<td>Advanced Digital Signal Processing</td>
<td>3.0</td>
</tr>
<tr>
<td>EENG515</td>
<td>Mathematical Methods for Signals and Systems</td>
<td>3.0</td>
</tr>
<tr>
<td>EENG519</td>
<td>Estimation Theory and Kalman Filtering</td>
<td>3.0</td>
</tr>
</tbody>
</table>

**ELECTIVES: Geological & Geophysics Foci**

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPGN533</td>
<td>Geophysical Data Integration &amp; Geostatistics</td>
<td>3.0</td>
</tr>
<tr>
<td>GPGN570</td>
<td>Applications of Satellite Remote Sensing</td>
<td>3.0</td>
</tr>
<tr>
<td>GEGN575</td>
<td>Applications of Geographic Information Systems</td>
<td>3.0</td>
</tr>
<tr>
<td>GEGN579</td>
<td>Python Scripting for Geographic Information Systems</td>
<td>3.0</td>
</tr>
<tr>
<td>GEOL557</td>
<td>Earth Resource Data Science 1: Fundamentals</td>
<td>3.0</td>
</tr>
<tr>
<td>GEOL558</td>
<td>Earth Resource Data Science 2: Applications and Machine-Learning</td>
<td>3.0</td>
</tr>
</tbody>
</table>

**ELECTIVES: Other Subject Area Foci**

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBBG525</td>
<td>Business Analytics</td>
<td>3.0</td>
</tr>
<tr>
<td>EBBG559</td>
<td>Supply Chain Analytics</td>
<td>3.0</td>
</tr>
</tbody>
</table>
**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 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 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:
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.

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).
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: CSCI101 or CSCI102 or CSCI128, MATH201 or MATH334, MATH332 OR Graduate level standing and at least CSCI128 or equivalent.

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. 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.

Professors
Douglas Nychka, Applied Mathematics & Statistics
Paul Sava, Geophysics
Michael Wakin, Electrical Engineering

Professor of Practice
Jim Crompton, Petroleum Engineering

Associate Professors
Soutir Bandyopadhyay, Applied Mathematics and Statistics
Donit Hammerling, Applied Mathematics & Statistics
Hua Wang, Computer Science

Teaching Associate Professor
Wendy Fisher, Computer Science

Research Associate Professor
Zane Jobe, Geology and Geological Engineering