DATA SCIENCE

The Master of Data Science (MDS) is a 12-month coursework program offered by the Department of Mathematics and Statistics that trains individuals to become computationally skilled and ethically minded data analysts. Students become well versed in key technologies in data science, including data wrangling, data mining, data integrity, visualization, machine learning, predictive modelling, and spatial-temporal methods. Through hands-on training, students analyze big data independently and collaboratively such that graduates are primed to help organizations translate data into knowledge and actionable insights. The program features in-class experiential learning opportunities, including how to address and describe complex problems relevant to industry partners, as well as how to explore ethical considerations of privacy, data security, objective analysis and visualization.

Administrative Staff

**Director and Graduate Program Coordinator**
Ayesha Ali (509 MacNaughton, Ext. 53896)
mdsdirector@uoguelph.ca (calendar.uoguelph.ca/graduate-calendar/graduate-programs/data-science/mdsdirector@uoguelph.ca)

**Graduate Program Assistant**
Susan McCormick (440 MacNaughton, Ext. 56553/52155)
mdsgrad@uoguelph.ca (calendar.uoguelph.ca/graduate-calendar/graduate-programs/data-science/mdsgrad@uoguelph.ca)

Graduate Faculty

**R. Ayesha Ali**
B.Sc. Western Ontario, M.Sc. Toronto, PhD Washington - Associate Professor
Graduate Faculty

**Daniel A. Ashlock**
B.Sc. Kansas, PhD CalTech - Professor and Chair
Graduate Faculty

**Neil Bruce**
B.Sc. Guelph, M.A.Sc., Waterloo, PhD York - Associate Professor
Graduate Faculty

**Monica Cojocaru**
BA, M.Sc. Bucharest, PhD Queen's - Professor
Graduate Faculty

**Rozita Dara**
B.Sc. Shahid Teheshti, M.Sc. Guelph, PhD Waterloo - Associate Professor
Graduate Faculty

**Lorna Deeth**
B.Sc., M.Sc., PhD Guelph - Assistant Professor
Graduate Faculty

**Ali Dehghantanha**
BSE Azad, M.Sc., PhD Putra Malaysia - Assistant Professor
Graduate Faculty

**Hermann J. Eberl**
Dipl. Math (M.Sc.), PhD Munich Univ. of Tech. - Professor

Graduate Faculty

**Zeny Feng**
B.Sc. York, MMath., PhD Waterloo - Professor
Graduate Faculty

**Dan Gillis**
B.Sc., M.Sc., PhD Guelph - Associate Professor
Graduate Faculty

**Andrew Hamilton-Wright**
B.Sc., M.Sc. Guelph, PhD Waterloo - Associate Professor
Graduate Faculty

**Julie Horrocks**
B.Sc. Mount Allison, BFA Nova Scotia College of Art & Design, MMath, PhD Waterloo - Professor
Graduate Faculty

**David Kribs**
B.Sc. Western, MMath, PhD Waterloo - Professor
Graduate Faculty

**Anna T. Lawniczak**
M.Sc. Wroclaw, PhD Southern Illinois - Professor
Graduate Faculty

**Xiaodong Lin**
BASc Nanjing, M.Sc. East China Normal, PhD Beijing, PhD Waterloo - Professor
Graduate Faculty

**Khurram Nadeem**
B.Sc., M.Sc. Karachi, PhD Alberta - Assistant Professor
Graduate Faculty

**Mihai Nica**
B.Math., Waterloo, PhD Courant Institute NYU - Assistant Professor
Graduate Faculty

**Stacey Scott**
B.Sc. Dalhousie, PhD Calgary - Professor
Graduate Faculty

**Fei Song**
B.Sc. Jilin (China), M.Sc. Academia Sinica (China), PhD Waterloo - Associate Professor
Graduate Faculty

**Fangju Wang**
BE Changsha, M.Sc. Peking, PhD Waterloo - Professor
Graduate Faculty

**Yang Xiang**
B.Ss., M.Sc. BUAA (Beijing), PhD British Columbia - Professor
Graduate Faculty

MDS Program

Admission Requirements

Upon recommendation by the Department of Mathematics and Statistics, admission to the Master of Data Science may be granted to applicants who have completed an honor’s Bachelor’s degree or equivalent from an accredited institution with a minimum overall average of 70% (B-) in the last four semesters of study with:
1) a major or minor in data science, computer science, mathematics, statistics, or a related field; or

2) working knowledge of statistics and computer programming, as demonstrated through completion of university or college level degree credit courses equivalent to the U of G courses STAT*3240 Applied Regression Analysis and CIS*2500 Intermediate Programming.

Please note: prospective students with an Honour’s Bachelor’s degree in an unrelated field who do not meet the above requirements may gain entry to the program after completing the Diploma in Applied Statistics (calendar.uoguelph.ca/undergraduate-calendar/special-study-opportunities/certificates-diplomas/applied-statistics/) (or equivalent) with a minimum overall average of at least 70% (B-).

Successful applicants must also meet the University of Guelph’s English Proficiency requirements for admission. If an applicant’s first language is not English, an English Language Proficiency test will be required during the application phase.

All applications will be received and reviewed by the Data Science Program Committee. The program especially encourages applications from qualified members of under-represented groups, particularly from those who self-identify as women, visible minorities and Indigenous peoples.

Program Requirements

Students in the Master of Data Science program are required to complete a minimum of 4.00 graduate credits, consisting of four core courses (2.00 credits), two electives (1.00 credits), and either the two capstone courses or DATA*6700 Data Science Project (1.00 credits).

Core Courses:

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA*6100</td>
<td>Introduction to Data Science</td>
<td>0.50</td>
</tr>
<tr>
<td>DATA*6200</td>
<td>Data Manipulation and Visualization</td>
<td>0.50</td>
</tr>
<tr>
<td>DATA*6300</td>
<td>Analysis of Big Data</td>
<td>0.50</td>
</tr>
<tr>
<td>DATA*6400</td>
<td>Machine Learning for Sequential Data Processing</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Capstone Courses:

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA*6500</td>
<td>Analysis of Spatial-Temporal Data</td>
<td>0.50</td>
</tr>
<tr>
<td>DATA*6600</td>
<td>Applications of Data Science</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Electives:

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIS*6020</td>
<td>Artificial Intelligence</td>
<td>0.50</td>
</tr>
<tr>
<td>CIS*6050</td>
<td>Neural Networks</td>
<td>0.50</td>
</tr>
<tr>
<td>CIS*6070</td>
<td>Discrete Optimization</td>
<td>0.50</td>
</tr>
<tr>
<td>CIS*6160</td>
<td>Multiagent Systems</td>
<td>0.50</td>
</tr>
<tr>
<td>CIS*6320</td>
<td>Image Processing Algorithms and Applications</td>
<td>0.50</td>
</tr>
<tr>
<td>ENGG*6070</td>
<td>Medical Imaging</td>
<td>0.50</td>
</tr>
<tr>
<td>ENGG*6100</td>
<td>Machine Vision</td>
<td>0.50</td>
</tr>
<tr>
<td>ENGG*6140</td>
<td>Optimization Techniques for Engineering</td>
<td>0.50</td>
</tr>
<tr>
<td>ENGG*6400</td>
<td>Mobile Devices Application Development</td>
<td>0.50</td>
</tr>
<tr>
<td>MATH*6020</td>
<td>Scientific Computing</td>
<td>0.50</td>
</tr>
<tr>
<td>MATH*6021</td>
<td>Optimization I</td>
<td>0.50</td>
</tr>
<tr>
<td>MATH*6022</td>
<td>Optimization II</td>
<td>0.50</td>
</tr>
</tbody>
</table>

MATH*6051 | Mathematical Modelling                     | 0.50    |
MATH*6071 | Biomathematics                              | 0.50    |
PHIL*6400 | Ethics of Data Science                     | 0.50    |
PLNT*6500 | Applied Bioinformatics                     | 0.50    |
STAT*6550 | Computational Statistics                   | 0.50    |
STAT*6801 | Statistical Learning                       | 0.50    |
STAT*6802 | Generalized Linear Models and Extensions   | 0.50    |
STAT*6721 | Stochastic Modelling                       | 0.50    |
STAT*6821 | Multivariate Analysis                      | 0.50    |
STAT*6841 | Computational Statistical Inference        | 0.50    |
STAT*6950 | Statistical Methods for the Life Sciences  | 0.50    |

DATA*6100 | Introduction to Data Science Fall Only [0.50]
This course includes an introduction to the methods of modern statistics such as splines, general additive models, principal components analysis, and classifiers. Students learn resampling methods such as bootstrap, cross-validation, boosting, and bagging. Methods of model selection include search-and-score and regularization, and students practice communicating technical ideas to a non-technical audience, including via data visualization.

Department(s): Department of Mathematics and Statistics
Location(s): Guelph

DATA*6200 | Data Manipulation and Visualization Fall Only [0.50]
This course provides a hands-on introduction to the manipulation and visualization of complex data sets using a programming language. Efficient techniques for importing and exporting data in various formats, data acquisition, data integrity, and good analysis practices are discussed. Several programming tools and libraries are introduced to restructure, transform and fuse disparate data types for visualization and data summaries in table format. Basics of manipulating space-time data is also covered.

Restriction(s): Restricted to Master of Data Science students.
Department(s): Department of Mathematics and Statistics
Location(s): Guelph

DATA*6300 | Analysis of Big Data Unspecified [0.50]
This course introduces software tools and data science techniques for analyzing big data. It covers big data principles, state-of-the-art methodologies for large data management and analysis, and their applications to real-world problems. Modern and traditional machine learning techniques and data mining methods are discussed and ethical implications of big data analysis are examined. May be offered in conjunction with CIS*6180.

Restriction(s): Credit may be obtained for only one of CIS*6180 or DATA*6300.
Department(s): School of Computer Science
Location(s): Guelph

DATA*6400 | Machine Learning for Sequential Data Processing Unspecified [0.50]
This course emphasizes machine learning for sequential data processing. It covers common challenges and pre-processing techniques for sequential data such as text, biological sequences, and time series data. Students are exposed to machine learning techniques, including classical methods and more recent deep learning models, so that they obtain the background and skills needed to confront real-world applications of sequential data processing. May be offered in conjunction with CIS*6190.

Restriction(s): Credit may be obtained for only one of CIS*6190 or DATA*6400.
Department(s): School of Computer Science
Location(s): Guelph
DATA*6500 Analysis of Spatial-Temporal Data Summer Only [0.50]
This course introduces software tools and data science techniques for analyzing big geospatial data. An overview of raster-based geographic information systems (GIS) for identifying patterns and clusters in spatial-temporal data using state-of-the-art software and programming languages is provided. Concepts such as kriging/Gaussian processes, vgrams and autoregressive correlation structures are discussed. Data summaries and visualizations specific to spatial-temporal problems are introduced.

**Restriction(s):** Restricted Master of Data Science students.
**Department(s):** Department of Mathematics and Statistics

DATA*6600 Applications of Data Science Summer Only [0.50]
This interdisciplinary team-taught seminar course provides students the opportunity to synthesize information, research methods, and present cutting-edge applications of data science. Learning outcomes include identifying reliable sources, understanding and presenting relevant contemporary data science methods, thinking critically about practical implementations of data science, and effective peer collaboration. Emphasis is placed on effectively communicating technical content and insights to a non-technical audience.

**Prerequisite(s):** DATA*6200 and DATA*6300
**Restriction(s):** Restricted to Master of Data Science students.
**Department(s):** Department of Mathematics and Statistics
**Location(s):** Guelph

DATA*6700 Data Science Project Unspecified [1.00]
This course is a one-semester research project course for students in the Master of Data Science program. In this course, students plan, develop, and write a faculty- or industry-led research paper, as well as present on their work. The project should advance knowledge or practice in data science or a closely related area, and address a real-world problem faced by industry. The project should focus on data science in the spatial and temporal dimension(s), to be approved by the course instructor.

**Restriction(s):** Instructor consent required.
**Department(s):** Department of Mathematics and Statistics
**Location(s):** Guelph