



DATA*6600 Applications of Data Science

Semester Year

0 COURSE PREAMBLE

Please note that the ongoing COVID-19 pandemic may necessitate a revision of the format of course offerings and academic schedules. Any such changes will be announced via Counselink and/or class email. This includes on-campus scheduling during the semester, midterms and final examination schedules. All University-wide decisions will be posted on the COVID-19 website and circulated by email.

Mental Health. One out of every five students in Canada experiences some sort of mental health issue at some point in their academic career. If you find yourself facing a mental health crisis, or just need to talk to someone, please consider taking advantage of one of the following resources available to University of Guelph students:

Counseling Services: (x53244) is located at Health Services (J.T. Powell Building) and offers individual and group counselling sessions by appointment or walk-in.

Student Support Network is located in Raithby House (across from the cannon) and offers confidential, peer-based, drop-in support.

Good2Talk: (1-866-925-5454) is a free, 24/7 student hotline that provides professional counselling and referrals for mental health, addictions and well-being.

Here 24/7: (1-844-437-3247) specializes in assessment, referral and appointment booking and is available 24/7 for crisis support.

You are not alone and you will not be judged for asking for help.

For information on current safety protocols, follow these links:

<https://news.uoguelph.ca/return-to-campuses/how-u-of-g-is-preparing-for-your-safe-return/>
<https://news.uoguelph.ca/return-to-campuses/spaces/\#ClassroomSpaces>

Please note, these guidelines may be updated as required in response to evolving University, Public Health or government directives. Medical notes will not normally be required for singular instances of academic consideration, although students may be required to provide supporting documentation for multiple missed assessments or when involving a large part of a course (e.g. final exam or major assignment).

1 INSTRUCTOR

This course is team-taught by graduate faculty from Department of Mathematics and Statistics, and the School of Computer Science at the University of Guelph. In its first iteration, the course will be taught by the following team:

Instructor: Ayesha Ali
Office: MACN 509
Phone extension: 53896
E-mail: aali@uoguelph.ca

Instructor: Alysha Cooper
Office: MACN
Phone extension:
E-mail: acoope05@uoguelph.ca

Instructor: Rozita Dara
Office: Reynolds 3311
Phone extension: 58762
E-mail: drozita@uoguelph.ca

2 AIMS & OBJECTIVES

2.1 Calendar Description

This interdisciplinary team-taught seminar 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 effective communication to a non-technical audience.

2.2 Course Description

Data Science is constantly evolving as a field and requires technically literate individuals that are effective communicators and critical thinkers. This course aims to refine students' skills in

these areas while exposing them to cutting-edge applications of data science. As a team-taught course, the course is divided into 4 short (equally-weighted) modules, students will see data science from the perspectives of faculty in Mathematics, Statistics and/or Computer Science. Students will be expected to work in groups, as well as read, critique and present current articles related to data science. Each group may also be required to present their own solutions to problems presented through case studies.

Prerequisite(s): DATA*6200 and DATA*6300

Restriction(s): Restricted to Master of Data Science students

Academic Department: Mathematics & Statistics

Credit Weight: 0.5

2.3 Learning Outcomes

Upon successful completion of this course, students will have demonstrated the ability to:

1. Identify and discuss various data science applications (such as genetic programming, cybersecurity, informed policymaking, to name a few) and potential real-world problems in these fields;
2. Propose multiple methodological approaches to resolving data-science related problems;
3. Collaborate with peers and employ a multidisciplinary approach to developing design solutions to data science related problems;
4. Consider, question, and critique alternative solutions presented by peer groups in consideration of technical, social, and ethical themes;
5. Clearly present proposed solutions to data science-related problems through written and oral forms of communication; and
6. Reflect upon and discuss ethical and social implications of data science applications.

2.4 Instructor's Role and Responsibility to Students

The role of each instructor is to facilitate discussion and provide feedback to students in each of their respective modules.

3 TEACHING AND LEARNING ACTIVITIES

3.1 Timetable

Lectures: Tu, Th at 10-11:20am EST Room: MCKN 318

Office hours: TBD and provided on Counselink once set. Office hours may change.

3.2 Course Topics and Schedule

Week	Topic
Week 1	Introduction
Weeks 2-3	Module 1: Communication and Reflection, Case Study A
Weeks 4-6	Module 2: Cybersecurity, Case Study B
Weeks 7-9	Module 3: Interactive Visualization, Case Study C
Weeks 10-12	Module 4: Causal Inference, Case Study D

4 LEARNING RESOURCES

4.1 Course Website

Course material, news, announcements, and grades will be regularly posted to the DATA*6600 Courselink site. You are responsible for checking the site regularly.

4.2 Required Resources

The course will use multiple resources* (textbooks, articles, and websites), including:

Suzan Last, *Technical Writing Essentials*, University of Victoria, 2019.

Claus O. Wilke, *Fundamentals of Data Visualization*, O'Reilly Media Inc., 2019.

Cynthia Dwork and Aaron Roth, [The Algorithmic Foundations of Differential Privacy](#), Foundations and Trends in Theoretical Computer Science, 2014.

*Additional resources or readings will be provided by instructors as needed.

5 ASSESSMENT

5.1 Dates and Distribution

Assignment	Due Date	Weighting	Learning Outcome(s) Assessed
Communication and Reflection: Self Reflection Assignment	Week 1	5%	1, 5, 6
Case Study A: Group Presentation*	Week 3	20%	2-6
Case Study B: Group Presentation*	Week 6	25%	1-6
Case Study C: Group Presentation*	Week 9	25%	1-6
Case Study D: Group Presentation*	Week 12	25%	1-6

*Based on class size, some assessments may not be done in groups. When individuals work in groups, their grade will be based on the group mark and peer-reviews by their fellow group members. Course instructors may also introduce an explicit part of the assignment that is completed individual.

5.2 Module Descriptions

Module 1: Communication and Reflection

Instructor: Ayesha Ali

This module will prompt students to critically assess their own ability to communicate effectively and to refine these soft skills that are essential for today's data scientist. Class discussions will involve critiquing samples of written and oral presentations of technical analyses. By the end of this module, students should be able to self-critique their own written and oral work and demonstrate proficiency in synthesizing technical information into a coherent presentation aimed at a non-technical audience.

Module 2: Data Science Applications in Cybersecurity

Instructor: Rozita Dara

This module reviews two aspects of data science in cybersecurity, which together comprise case study B: 1) differential privacy and its applications in protecting individuals' data; and 2) applications of data science in data security. Balancing ethics of privacy, transparency and utility will be discussed throughout.

Case study B1: Differential privacy is an effective approach for protecting individuals' privacy while preserving data utility for analytics purposes. The first part of this case study, students will examine methods and software tools that can be used to anonymize and de-identify sensitive data in applications such as healthcare and finance. Group presentations and class discussions will focus on differential privacy techniques (e.g. machine learning and statistical), applications such as answering distributed queries, and examining privacy and data utility trade-offs.

Case study B2: The second part of this case study will focus on applications of machine learning and statistical modeling in data security will be discussed. Students will work in groups to investigate a security application such as intrusion detection and present approaches that can be used to detect threats and cyber attacks. Presentations will review the data sources, different methods that the problem can be formulated, potential algorithms that can be used to tackle threat identification tasks, methods for evaluation of the approach, and best practices for reporting the results.

Module 3: Interactive Visualizations and Dashboard Design

Instructor: Ayesha Ali

This module will look at best practices for dashboard design, including ethical considerations, and students will become familiarized with the Shiny app in R for producing effective dashboards. Dashboards can provide a one-stop-shop experience for synthesizing data from multiple sources and creating visualizations that help the reader monitor, analyze, and display key metrics to answer specific questions. They form a standard tool in the visualization toolkit of today's data scientist. By the end of the module student should be able to build a dashboard and explain their rationale for the design choices made and the statistical or machine learning methods used to analyze the data and produce the visualizations.

Module 4: Causal Inference

Instructor: Alysha Cooper

This module will focus on causal inference – methods for inferring causal relations from data. Causal inference has its roots in psychology, econometrics, epidemiology and more recently in artificial intelligence. Students will review the concepts of cause and effect within a statistical framework, including concepts such as confounding, Simpson's paradox, interventions, causal effects (direct/indirect) and causal mediators. Depending on time and interest, topics may include causal graphs, propensity score matching, inverse-probability-of-treatment weighting, structural equation modelling, and marginal structural models. Relevant software tools will be introduced as needed.

5.3 Course Grading Policies

Accommodation of Religious Obligations: If you are unable to meet an in-course requirement due to religious obligations, please email the course instructor within two weeks of the start of the semester to make alternate arrangements. See the graduate calendar for information on regulations and procedures for Academic Accommodation of Religious Obligations:

https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/sec_d0e2228.shtml

Passing grade: In order to pass the course, students must obtain a grade of 65% or higher on the total mark of all assessments.

6 UNIVERSITY STATEMENTS

6.1 Email Communication

As per university regulations, all students are required to check their e-mail account regularly; e-mail is the official route of communication between the University and its students.

6.2 When You Cannot Meet a Course Requirement

If, due to medical, psychological, or compassionate circumstances you find yourself unable to complete an in-course requirement, please advise the course instructor (or designated person, such as a teaching assistant) in writing, with your name, id#, and e-mail contact. The grounds for Academic Consideration are detailed in the Undergraduate and Graduate Calendars.

Undergraduate Calendar - Academic Consideration and Appeals

[<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-ac.shtml>]

Graduate Calendar - Grounds for Academic Consideration

[https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/sec_d0e2182.shtml]

6.3 Drop Date

Courses that are one semester long must be dropped by the end of the last class day; two-semester courses must be dropped by the last day of class in the second semester. The regulations and procedures for changing graduate course registration are available in the Undergraduate and Graduate Calendars.

Undergraduate Calendar – Dropping Courses

[<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-drop.shtml>]

Graduate Calendar – Registration Changes

[<https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/genreg-reg-regchg.shtml>]

6.4 Copies of Out-of-class Assignments

Keep paper and/or other reliable back-up copies of all out-of-class assignments; you may be asked to resubmit work at any time.

6.5 Accessibility

The University promotes the full participation of students who experience disabilities in their academic programs. To that end, the provision of academic accommodation is a shared responsibility between the University and the student.

When accommodations are needed, the student is required to first register with Student Accessibility Services (SAS). Documentation to substantiate the existence of a disability is required; however, interim accommodations may be possible while that process is underway.

Accommodations are available for both permanent and temporary disabilities. It should be noted that common illnesses such as a cold or the flu do not constitute a disability.

Use of the SAS Exam Centre requires students to book their exams at least seven days in advance and not later than the 40th class day.

More information can be found on the SAS website [<https://www.uoguelph.ca/sas>]

6.6 Academic Integrity

The University of Guelph is committed to upholding the highest standards of academic integrity, and it is the responsibility of all members of the University community—faculty, staff, and students—to be aware of what constitutes academic misconduct and to do as much as possible to prevent academic offences from occurring. University of Guelph students have the responsibility of abiding by the University's policy on academic misconduct regardless of their location of study; faculty, staff, and students have the responsibility of supporting an environment that discourages misconduct. Students need to remain aware that instructors have access to and the right to use electronic and other means of detection.

Please note: Whether or not a student intended to commit academic misconduct is not relevant for a finding of guilt. Hurried or careless submission of assignments does not excuse students from responsibility for verifying the academic integrity of their work before submitting it. Students who are in any doubt as to whether an action on their part could be construed as an academic offence should consult with a faculty member or faculty advisor.

Undergraduate Calendar – Academic Misconduct

[<https://www.uoguelph.ca/registrar/calendars/undergraduate/current/c08/c08-amisconduct.shtml>]

Graduate Calendar – Academic Misconduct

[https://www.uoguelph.ca/registrar/calendars/graduate/2018-2019/genreg/sec_d0e2632.shtml]

6.7 Recording of Materials

Presentations that are made in relation to course work—including lectures—cannot be recorded or copied without the permission of the presenter, whether the instructor, a student, or guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

6.8 Resources

The Academic Calendars [<https://www.uoguelph.ca/academics/calendars>] are the source of information about the University of Guelph's procedures, policies, and regulations that apply to undergraduate, graduate, and diploma programs.