Dalhousie University

Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq. We are all Treaty people.[1]

Faculty of Computer Science / Department of Community Health and Epidemiology (Faculty of Medicine)

Course Proposal
CSCI6XXX/4147/CH&E6XXX: Applied Research in Health Data Science

Summer (A) 2022

M/W/F: 1300-1500, 1201 Mona Campbell Building

COURSE INFORMATION

Instructor Information

- **Instructor:** Finlay Maguire
- **Office:** 4239 Mona Campbell Building, Studley Campus
- **Office Phone:** TBD
- **Office Hours:** Wednesday 1500-1700 (following synchronous session)
- **Email:** finlay.maguire@dal.ca

Course Description (University Calendar)

This course is an introduction to the application of data science methods to health data within interdisciplinary research contexts. Students will be introduced to the main types of health data and their principal analysis methods while developing key research skills specific to effectively working at the intersection of medicine and computer science. This will encompass developing technical skills in the robust/reproducible analysis of data from medical databases, radiological imaging, electronic medical records, and physiological time-series data. Students will also gain specific training in developing interdisciplinary health data science research proposals including key considerations such as research ethics, data legislation, knowledge translation, and effective collaboration.

[1] For more information about the purpose of territorial acknowledgements, or information about alternative territorial acknowledgements if your class is offered outside of Nova Scotia, please visit https://native-land.ca/.

Course Description (Syllabus)

Health data science is a rapidly growing research field across academia, government, and industry. It relates to the application of statistical and machine learning approaches to analyse large complex medical datasets including electronic medical records, radiological imaging, physiological sensor data, and longitudinal databases. This course combines an overview of these key types of medical data, hands-on introduction to their principal analysis methods, and training in how to apply them in interdisciplinary research contexts. Using a combination of lectures, R-based exercises, student-driven tutorials, and collaborative development of a research proposal, students will gain the skills necessary to plan and conduct effective health data science research.

Class Format

This class is formatted around a mixture of didactic lectures (Mondays), assessed in-class practical exercises (Fridays), and tutorials around primary literature (Wednesday 1st half). These tutorials will involve a rotation of students presenting papers and class-based discussion of the strengths/weaknesses and key methodological take-aways of the presented work. Additionally, the main assessment output for the course will be a collaboratively developed research proposal supported by in-class time and proposal/research-skills related tutorials (Wednesday 2nd half).

Minimal Technical Requirements

This course will require access to an internet-enabled computer capable of installing and running Rstudio (>1.0). Rstudio is freely available and can be install from here: https://www.rstudio.com/products/rstudio/

Course Pre-requisites, Co-requisites, Exclusions and/or other Restrictions

Students should either have some previous programming experience (ideally with R) and a knowledge of basic machine learning and/or statistical methods. For graduate students this will be self-certified and relevant additional training material can be found in course readings. For undergraduate students pre-requisites of CSCI2110 and either STAT2060 or CSCI2360 are required.

Basic pre-course familiarisation with R is recommended for all students e.g., completion of the Harvard-Chan School Bioinformatics training module: https://hbctraining.github.io/Training-modules/IntroR/

Course Rationale and/or Other Restrictions and Requirements

This course is designed as an elective course for graduate students in Community Health & Epidemiology and graduate/advanced undergraduate students in the Faculty of Computer Science interested in working at the intersection of medicine and computer science. The primary goals of this course are to (i) provide an overview of the main types of medical data, (ii) introduce key analysis methods for each data, and (iii) build skills necessary for effective interdisciplinary research in this area. This will complement existing non-cross listed/co-located CS and CH&E courses by providing students with an introduction to a wide-range of concepts across those courses or an opportunity to apply those skills within a growing interdisciplinary research context. Within FCS, it will complement existing courses focused on specific analysis methods (e.g., CSCI 6504/6505/6509/6515/6612) and research skills focused courses (e.g., CSCI 6055/6061). Similarly, within CH&E, it will complement technical-skill focused courses (e.g., CH&E 6054/6056) as well as research training (e.g., HINF 6020.03/CH&E 8040) by supporting specific training at their intersection. Given the topic, goals, and cross-listing/co-locating of this course, it will be well placed to form part of the Master's of Digital Innovation.

Course Learning Outcomes

The aim of this course is to provide students with the skills and knowledge required to plan effective research in the application of data science approaches to medical data. Specifically, by the end of the course students will:

1. Understand the 4 principal sources and data types of medical data: longitudinal databases (tabular), electronic medical records (structured, semi-structured, and unstructured text), radiological imaging (image), and physiological (signal and time-series).
2. Identify and apply appropriate data science methods to the analysis of each data type
3. Gain the technical skills necessary for effective health data science research including data management, reproducibility, and version control.
4. Understand the key collaborative, legal, ethical, and knowledge translation concepts required in interdisciplinary health data science research.
5. Critically appraise research literature in health data science.
6. Combine these skills to develop high-quality collaborative health data science research proposals

Required Text(s)

- R for Data Science by Wickham & Grolemund (freely available from authors: https://r4ds.had.co.nz/)
- Hands on Machine learning for R by Boehmke & Greenwell (freely available from authors: https://bradleyboehmke.github.io/HOML/)
- Text Mining with R by Julia Slige & Davin Robinson (freely available from authors: https://www.tidytextmining.com/)

Important Dates

- Classes begin: **May 9th**
- Last Day to Change and Add Classes for registered students: **May 17th**
- Victoria Day - University Closed: **May 23rd**
- Last Day to Drop without "W": **May 26th**
- Last Day to Change from Audit to Credit and Vice Versa: **May 26th**
- Last Day to Drop with "W": **June 13th**
- Classes end: **June 27th**

Tentative Course Schedule

Week	Day	Topic	Format
1	Mon	Reproducible Research	**Lecture:** Introduction to Health Data Science
	Wed		**Tutorial**: 2 papers: reproducible science theory vs practice **Proposal Class:** Introduction to proposal and group formation
	Fri		**Practical**: Use of git, notebooks, (formative non-graded)
2	Mon	Medical Databases	**Lecture:** Medical Databases (Ontology codes, survey

			weights, manifold learning, boosted decision trees)
	Wed		**Tutorial:** 2 papers: manifold learning and boosted decision trees. **Proposal Class:** Data access considerations
	Fri		**Practical:** Using medical databases
3	Mon	Electronic Medical Record	**Lecture:** Electronic Medical Records (Structured, semi-structured, unstructured text, NLP, Language Models)
	Wed		**Tutorial:** 2 papers: EMR+NLP and Ontology based models **Proposal Class:** Research Ethics
	Fri		**Practical:** Using EMR data
4	Mon	Radiological Imaging	**Lecture**: Medical imaging, data formats, computer vision classical vs deep learning.
	Wed		**Tutorial:** 2 papers: pathology slide scoring, MRI **Proposal Class:** Intermediate check-in.
	Fri		**Practical:** Automating scoring of cancer risk
5	Mon	Physiological Signal Data	**Lecture:** Signal data and time-series, transformation, LSTMs
	Wed		**Tutorial:** 2 papers: wearable health monitors and EEG **Proposal:** Knowledge Translation
	Fri		**Practical:** Seizure prediction EEG
6	Mon	Proposal	**Proposal Presentations**

	Wed		**Proposal Presentations**
	Fri		**Proposal Presentations**

Course Assessments

Assessment	**Date of Evaluation**	**Weight**
Practical 1: Medical Databases	May 27th	10%
Practical 2: Electronic Medical Records	June 3rd	10%
Practical 3: Radiological Imaging	June 10th	10%
Practical 4: Physiology Time-Series	June 17th	10%
Proposal Presentation	June 20th	20%
Proposal Write-Up	June 27th	20%
Tutorial Paper Presentation	Throughout	10%
Participation	Throughout	10%

The first practical on "Health Data Science" tools will not be graded but attendance is required.

Grading: This course uses the standard Dalhousie grading scheme: (https://www.dal.ca/campus_life/academic-support/grades-and-student-records/grade-scale-and-definitions.html) Note**:** for graduate students the minimum passing grade is B- so any mark <70% will be converted to an F

Submission: Assignments must be submitted via Brightspace.

Late Policy: Late assignments will be penalised at 20% per day. Assignments submitted more than 5 days late can still be evaluated for feedback but the final grade will be 0. Late submissions without penalty will be considered **only if** the appropriate channels (student declaration of absence, doctor's note) are used.

Undergraduate Criteria: will be assessed using the same criteria as graduate students with the exception of having their worst practical assignment score dropped (resulting in 13.33% weighting for remaining 3 practical assignments). Undergraduates can also receive a minimum passing grade of C- compared to B- for graduate students.

Evaluation Criteria

Practical Assignments: each of the practical assignments draws on material from the related module of the course and applies those methods and to problem data sets. This will start with a worked step-by-step example data set (introduced and explained in the practical session by the instructor) with short knowledge and understanding questions. Following this, students will then complete a short report applying the full analysis approach to a new dataset. These assignments are to be completed during the practical session or outside of class time before the next practical (1 week's time). Collaboration is **not** permitted on these assignments.

Tutorials: students will be assigned a relevant paper from the health data science literature and will present the paper to the class and lead subsequent discussion and appraisal of the methods and results of the research. Presentation assignment will take place on the first day of class.

Grade	Criteria
100	Demonstrates excellent preparation and comprehension of the papers presented. Able to critically appraise the material, identify strengths and limitations and suggest alternative approaches. Can lead the discussion independently and responds to other students' comments thoughtfully
80	Demonstrates good preparation and comprehension of the papers presented. Able to critically appraise most of the material and identify key strengths and limitations. Contributes actively to the discussion.
60	Demonstrates adequate preparation and comprehension of the papers presented. Offers straightforward information and some insight into key strengths and limitations. Contributes sporadically to the discussion
40	Demonstrates minimal preparation and very limited comprehension of the papers presented.Offers only straightforward information and has no understanding of strengths and limitations. Contributes only very infrequently to the discussion.

Participation: students will be expected to participate in the tutorial discussions and in-class.

Component	Very Good to Excellent (4-5)	Good (3)	Fair to Poor (≤2)	Score
Preparation	Arrives fully prepared.	Arrives mostly, though not full, prepared.	Arrives with limited preparation.	/5
Engagement	Actively supports, engages, and listens to peers.	Makes a sincere effort to interact with peers.	Limited interaction with peers.	/5
Initiative	Plays an active role in discussions or actively asks questions.	Participates constructively in discussions or lectures.	Follows the discussion groups or the class.	/5
Quality	Questions or comments reflect analysis,	Questions or comments are relevant to the topic	Questions or comments are generally vague or too far	/5

	synthesis, and evaluation; level of discussion is consistently better because of the student's presence.	discussed; level of discussion is occasionally better because of the student's presence.	from the topic discussed; level of discussion is not affected by the student's presence.	
Total				/20

Research Proposal: students will be assigned to groups in the first week of class and throughout the course will collaboratively develop a research proposal. There will be weekly class time dedicated to training in the specifics of developing research proposals and for group work on the proposal. However, the groups will be expected to coordinate with one another and work on the proposal outside of class time as well. The final reports will be presented by the groups to the rest of the class during the last week of the course with Q&A from the other course participants. This will be followed by submission of a group written report including a statement of contribution that clearly indicates each team member's contribution to the proposal. The follow rubric summarises how this will be graded:

Component	**Very Good to Excellent** **4-5/5** **7-10/10**	**Good** **3/5** **4-6/10**	**Fair to Poor** **0-2/5** **0-3/10**	**Score**
Teamwork	Contributions statement and presentation show signs of strong equitable teamwork. Clear indication of a high level of mutual respect and collaboration within the group.	Contributions statement and presentation show signs of generally strong and equitable teamwork. Indication of a good level of mutual respect and collaboration within the group.	Contributions statement and presentation show little indication of strong or equitable teamwork. Signs of low levels of mutual respect and collaboration within the group.	/5
Research Question	Engaging and clearly defined interdisciplinary research question that solves an important extant issue in the field of health data science.	Relatively well defined interdisciplinary research question that addresses a relatively important extant issue in the field of health data science.	Poorly defined question with limitations in scope, importance, or interdisciplinarity.	/5
Abstract	A concise, informative, and well-written summary of the proposal targeted to an expert audience.	A relatively well written summary of the proposal targeted to an expert audience.	A poorly written summary that only conveys limited details about the proposal.	/5
Lay Summary	A concise, well-written plain-language summary	A generally clear plain-language summary	Poorly written plain-language summary	/5

	aimed at the general public summary which articulates the problem, proposed solution, and relevance in an accessible and clear manner.	aimed at the general public summary which largely articulates the problem, proposed solution, and relevance in a mostly accessible way.	that doesn't convey the problem, proposed solution, or relevance in an accessible manner.	
Introduction	A detailed and well-written technical overview that positions the research idea within the wider research and social context, articulates the knowledge gap the project will address, and introduces relevant methods	Generally detailed and well-written technical overview that positions the research idea within the wider research and social context, articulates the knowledge gap the project will address, and introduces relevant methods	A detailed and well-written technical overview that positions the research idea within the wider research and social context, articulates the knowledge gap the project will address, and introduces relevant methods	/10
Literature Review	Critically appraises and summarises a broad sampling of relevant literature that fully justifies the research question and proposed methodology.	Critically appraises and summarises a reasonably broad sampling of relevant literature that mostly justifies the research question and proposed methodology.	Reports a limited range of relevant literature with limited critical appraisal and doesn't justify the research question and proposed methodology.	/10
Methodology	Clearly communicates a reasonable, appropriate, and fully-developed methodology (including data gathering/access) that solves the research question and is well supported by the literature review.	Communicates a reasonable, appropriate, and relatively developed methodology (including data gathering/access) that solves the research question and is reasonably well supported by the literature review.	Limited methodology with significant gaps in, disconnection from the key research question, and limited literature support.	/10
Ethics	Thoughtfully and robustly explores the research ethics, potential obstacles/risks, and the risk/benefits of the research question, methodology, and impact of the proposed work.	Relatively strong exploration of the research ethics, potential obstacles/risks, and the risk/benefits of the research question, methodology, and impact of the proposed work.	Limited treatment of the ethics, potential obstacles/risks, and the risk/benefits of the research question, methodology, and impact of the proposed work.	/10

Discussion	Clearly addresses potential limitations in the proposed project, identifies, explores potential future directions/extensions, and evaluates implications of a reasonable range of potential results.	Addresses most potential limitations in the proposed project, identifies some future directions/extensions, and explores some implications of a reasonable range of results.	Only addresses a small range of limitations, implications, and/or future research directions.	/10
Knowledge Translation	Develops a robust and impactful plan to effectively mobilise knowledge gained from the proposed research across a range of sectors/settings.	Develops a relatively complete plan to mobilise knowledge gained from the proposed research across a range of sectors/settings.	Develops a limited plan that doesn't successfully mobilise potential findings and/or does so across a limited range of sectors/settings.	/10
Formatting	Proposal is clearly formatted with consistent citations and use of section headings, figure legends, and appendices where appropriate.	Proposal is mostly clearly formatted with some minor formatting errors e.g., malformed citations, inconsistency, missing sections.	Poorly formatted proposal with numerous issues with use of sections, citations, and presentation of figures/tables.	/5
Presentation	Well crafted presentation, with good use of slides to clearly convey components of the proposal.	Reasonably clear presentation, that mostly conveys key components of the proposal.	Less well organised presentation that doesn't convey key components of the proposed work.	/10
Presentation: Q&A	Well-prepared team that comprehensively handles questions that arise about their work.	Well-prepared team that handles questions that arise about their work relatively well.	Relatively poor team-wide ability to answer questions about their proposed work.	/5
Total				/100

UNIVERSITY STATEMENTS

Territorial Acknowledgement:

Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq.

We are all Treaty people.[2]

Internationalization
At Dalhousie, "thinking and acting globally" enhances the quality and impact of education, supporting learning that is "interdisciplinary, cross-cultural, global in reach, and orientated toward solving problems that extend across national borders."

Academic Integrity
At Dalhousie University, we are guided in all of our work by the values of academic integrity: honesty, trust, fairness, responsibility and respect. As a student, you are required to demonstrate these values in all of the work you do. The University provides policies and procedures that every member of the university community is required to follow to ensure academic integrity.

Accessibility
The Student Accessibility Centre is Dalhousie's centre of expertise for matters related to student accessibility and accommodation.
If there are aspects of the design, instruction, and/or experiences within this course (online or in-person) that result in barriers to your inclusion please contact:

- the Student Accessibility Centre (for all courses offered by Dalhousie with the exception of Truro)
- the Student Success Centre in Truro for courses offered by the Faculty of Agriculture

Your classrooms may contain accessible furniture and equipment. It is important that these items remain in place, undisturbed, so that students who require their use will be able to fully participate.

Conduct in the Classroom – Culture of Respect
Substantial and constructive dialogue on challenging issues is an important part of academic inquiry and exchange. It requires willingness to listen and tolerance of opposing points of view. Consideration of individual differences and alternative viewpoints is required of all class members, towards each other, towards instructors, and towards guest speakers. While expressions of differing perspectives are welcome and encouraged, the words and language used should remain within acceptable bounds of civility and respect.

Diversity and Inclusion – Culture of Respect
Every person at Dalhousie has a right to be respected and safe. We believe inclusiveness is fundamental to education. We stand for equality. Dalhousie is strengthened in our diversity. We are a respectful and inclusive community. We are committed to being a place where everyone feels welcome and supported, which is why our Strategic Direction prioritizes fostering a culture of diversity and inclusiveness (Strategic Priority 5.2).

Code of Student Conduct
Everyone at Dalhousie is expected to treat others with dignity and respect. The Code of Student Conduct allows Dalhousie to take disciplinary action if students don't follow this community expectation. When appropriate, violations of the code can be resolved in a reasonable and informal manner—perhaps through a restorative justice process. If an informal resolution can't be reached, or would be inappropriate, procedures exist for formal dispute resolution.

---

[2] For more information about the purpose of territorial acknowledgements, or information about alternative territorial acknowledgements if your class is offered outside of Nova Scotia, please visit https://native-land.ca/.

Fair Dealing policy
The Dalhousie University Fair Dealing Policy provides guidance for the limited use of copyright protected material without the risk of infringement and without having to seek the permission of copyright owners. It is intended to provide a balance between the rights of creators and the rights of users at Dalhousie.

Originality Checking Software
The course instructor may use Dalhousie's approved originality checking software and Google to check the originality of any work submitted for credit, in accordance with the Student Submission of Assignments and Use of Originality Checking Software Policy. Students are free, without penalty of grade, to choose an alternative method of attesting to the authenticity of their work, and must inform the instructor no later than the last day to add/drop classes of their intent to choose an alternate method.

UNIVERSITY POLICIES, GUIDELINES, AND RESOURCES FOR SUPPORT
Dalhousie courses are governed by the academic rules and regulations set forth in the Academic Calendar and the Senate.

University Policies and Programs
- Important Dates in the Academic Year (including add/drop dates)
- Classroom Recording Protocol
- Dalhousie Grading Practices Policy
- Grade Appeal Process
- Sexualized Violence Policy
- Scent-Free Program

Learning and Support Resources
- Academic Support - Advising Halifax, Truro
- Student Health & Wellness Centre
- On Track (helps you transition into university, and supports you through your first year at Dalhousie and beyond)
- Indigenous Student Centre. See also: Indigenous Connection.
- Elders-in-Residence: The Elders in Residence program provides students with access to First Nations elders for guidance, counsel and support. Visit the office in the Indigenous Student Centre or contact the program at elders@dal.ca or 902-494-6803.
- Black Student Advising Centre
- International Centre
- South House Sexual and Gender Resource Centre
- LGBTQ2SIA+ Collaborative
- Dalhousie Libraries
- Copyright Office
- Dalhousie Student Advocacy Service (DSAS)
- Dalhousie Ombudsperson
- Human Rights & Equity Services
- Writing Centre
- Study Skills/Tutoring