

UNIVERSITY OF TORONTO

MIE 1626

Data Science Methods and Statistical Learning

Fall 2024

SYLLABUS

Instructor:

Prof. Samin Aref

E-mail: aref@mie.utoronto.ca (prefix subject with MIE1626)

Lectures:

Thursday mornings at [GB 304](#) (see Quercus for the time).

Practical Sessions (Python Labs and Support Sessions):

Fridays 17:00-19:00 at [AB 107](#).

Instructor's office hours (see course schedule for exceptions):

The office hour of Prof. Aref is online only (time and Zoom link are available on the Modules page of this course on Quercus). Students should register in an online queue using a Google form (on the Modules page) at most 1 hour before the start time to be then admitted to the Zoom room in the order of registration.

Teaching Assistants:

TA: Rahul Patel	rm.patel@mail.utoronto.ca	Role: Tutorials/Piazza/Project 1
TA: Shay Chavoshian	shay.chavoshian@mail.utoronto.ca	Role: Tutorials/Piazza/Project 2

Course Description:

This is an intensive course which requires active engagement and participation. It will equip the students with the fundamental skills and knowledge for:

- combining, pre-processing, and cleaning data in practical data science projects,
- performing exploratory data analysis and visualizing data using Python and its libraries,
- participating in intellectual debates on predictive modeling and data-driven analysis,
- analyzing data and making inference using methods from statistical learning,
- resampling and bootstrapping to evaluate the error of different computational estimates,
- using confidence intervals and statistical hypothesis testing to explain patterns in data,
- implementing and interpreting advanced regression models for high-dimensional data,
- analyzing networked data using centrality measures and community detection,
- describing LLMs and using them for feature-based adaptation and text classification,
- designing and understanding ensemble models and support vector machines,
- enhancing the interpretability of black-box models using ML explainability methods,
- conducting multiple tests while controlling family-wise error rate and false discovery rate,
- designing online controlled experiments for causal inference,
- reasoning quantitatively using Bayesian vs. frequentist approaches and Fermi's estimation,
- designing fair and justifiable predictive models using fundamental AI fairness concepts.

Course Goals:

- Understand data as the foundation of quantitative analysis and reasoning.
- Acquire additional methods for modeling and analyzing data and making inference.
- Understand the statistical foundations of data science and machine learning methods.
- Replace black boxes with explainable and justifiable predictive models.
- Learn to implement the above principles, methods, and techniques using Python, and its data science and machine learning libraries to a wide range of applications.
- Learn how to interpret the results of statistical analyses.
- Learn how to detect and avoid false narratives disguised as quantitative evidence or resulted from sloppy or incorrect analysis.
- Learn how to plan, execute, and deliver successful data science projects using scientifically sound methods.

Main Channel of Communication: Piazza

The preferred contact method for getting a quick response from the instructor and/or the teaching assistants is Piazza (using *Post to Instructor(s)* – select “instructors” to include everyone from the instruction team as recipients).

Communication via email (prefix subject with MIE1626) is fine if you have a reason for not using Piazza for that question. You can email any of the TAs if you have difficulty with starting to use the online tools of this course (Piazza, Quercus, etc.).

Questions in the general forms of “is this the correct answer?” or “what is wrong with my code?” or “why my code does not compile?” and the like will not receive a response.

Class Resources:

You will need to log into Quercus to gain access to course material, and obtain regular course information (e.g., downloading lecture materials, etc.), complete quizzes for reading assignments, receive grade/feedback, and access important announcements. This can be done with your UTORid at: <https://q.utoronto.ca/>. Students are expected and strongly recommended to fully participate in all lectures and practical sessions for the duration of the course. The preferred method for Python programming in this course is via [Google Colaboratory](#).

Ask all your questions in class, lab, office hours, or on Piazza and ask many. We love answering your questions as many as they are or as frequent as they are.

Textbooks:

Slides and recommended resources will be posted throughout the course. A large portion of the lectures will be based on the following textbook:

- “An Introduction to Statistical Learning with Applications in Python”, 3rd Edition, by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor, 2023, Springer (free [link](#))

Other suggested reading:

- “The Elements of Statistical Learning”, 2nd Ed, 12th printing, by Trevor Hastie et al., 2009, Springer (free [link](#))
- “The StatQuest Illustrated Guide to Machine Learning” by Josh Starmer, 2022
- “Fairness and Machine Learning - Limitations and Opportunities” by Solon Barocas, Moritz Hardt, and Arvind Narayanan, 2023, MIT Press (free [link](#))

Evaluation:

Assessment	Weight (%)	Schedule/Deadline
Eight reading assignment quizzes	8	Deadline for reading assignment one: Sep 17 th at 21:00 Deadlines for other reading assignments: As per course schedule
Project 1	16	As per course schedule
Midterm Assessment	20	As per course schedule
Project 2	16	As per course schedule
Final Assessment*	38	As per course schedule
Active Participation	2	During labs and lectures
Bonus for Piazza ⁺	(+2 points)	As per course schedule

*The Final Assessment is *mandatory* and will result in a grade of FZ (failing the course) assigned on the transcript if not completed.

+Top ten students, ranked on the number of endorsed answers on Piazza, will have 2 bonus points added to their final course grade. A minimum of 3 endorsed answers are required to qualify for the bonus.

Course Structure:

Synchronous Activities

- Weekly Lectures
- Weekly Practical sessions (either a Tutorial or a Q&A session as per course schedule)

Asynchronous Activities

- Two projects are to be completed individually throughout the semester (see *Evaluation*) and submitted via Quercus by their deadline.
- Eight quizzes associated with eight reading assignments are to be completed throughout the semester (see *Evaluation*) and submitted via Quercus by their deadline.
- All questions related to the course materials and project queries should be posted on Piazza. Please do not share any code/results/figures/solutions related to the projects on Piazza. Emailed questions will not be answered.
- You are encouraged to participate by asking (and answering) course questions. We will endeavour to respond to course inquiries as quickly as possible, but there is no guarantee that answers will be provided outside of business hours.

Penalty for late submissions (projects and reading summaries):

A submission is either *in time* (before the deadline) or *late* (after the deadline). Projects and reading assignment quizzes should be submitted before the deadline; ideally hours before the deadline to avoid any risk of late submission. Quercus submission time will be used. Late projects and reading assignment submission will incur an automatic penalty as follows:

- -30% (of project/task maximum mark) if submitted within 72 hours past the deadline.
- A mark of zero will be given if the submission is 72 hours late or more.

Re-grading:

If a student wishes to discuss marking for a project or the midterm assessment, they should first meet with a TA during the next available practical session (tutorial). If the issue remains unresolved, the instructor is to be contacted directly by email with the TA (the marker) in cc.

Reading Assignments:

Eight quizzes for reading assignments are required to be completed by each individual student for eight weeks of the course (one per week) (see *Evaluation*; see *Course Schedule*).

Reading assignment 1: Robert Matthews (2000) Storks deliver babies (p=0.008). Teaching Statistics 22:36-38 [link1](#) [link2](#)

Other reading assignments will be announced on Quercus.

Academic Integrity:

The University of Toronto expects you to be a full member of the academic community and to observe the rules and conventions of academic discourse. In particular, all the work you submit must be your own and no part of your submitted work should be prepared by someone else. Plagiarism or any other form of cheating in examinations, tests, assignments, or projects, is subject to serious academic penalty (e.g., suspension or expulsion from the faculty or university). Also, doing anything for the purpose of aiding or assisting another student to commit plagiarism is an offence which makes both parties liable for bearing the penalties and consequences. The full text of the policy that governs Academic Integrity at U of T (the Code of Behaviour on Academic Matters) can be found at:

www.governingcouncil.utoronto.ca/policies/behaveac.htm

Here are a couple of general guidelines to help you avoid plagiarism:

- Never look at another student's homework. Do not show another student your solution. This applies to all drafts of a solution and to incomplete and even incorrect solutions.
- Keep discussions with other students focused on *concepts* and *examples*. Any code or solutions that you submit should be yours alone.
- Do not post any of your assignment questions in a private or public online discussion forum, social media and messaging groups, or website in order to solicit solutions from others.

Note that, under the University of Toronto code of conduct, a person who supplies an assignment or project to be copied will be penalized in the same way as the one who makes the copy.

Use of Automated Plagiarism Detection Methods:

Several plagiarism detection tools will be used to assist in the evaluation of the originality of the submitted work for both text and code. They are quite sophisticated and difficult to defeat. Submissions go through automated plagiarism detection tools as a part of the evaluation.

Policy regarding the use of AI assistant tools:

If you use a generative AI tool (e.g., ChatGPT, GitHub Copilot, or any other similar tool) to get ideas and/or partial answers for projects or assignments or to generate any code and/or text, you must declare in your submission the tools that you have used and describe the usage, and include in your submission an appendix that captures all the interactions (prompts and responses). You will not be penalized for the declared use of such AI assistant tools, and the grading decisions will be based on your original contributions as well as the efforts that you make to check and/or correct the answers provided by the tool. Students are ultimately accountable for the work they submit. Failure to fully declare the use of this tool will be considered "unauthorized assistance" and will have consequences (See B.I. in the [U of T Code of Behaviour on Academic Matters](#)).

Suspected Plagiarism:

The following constitute plagiarism on project and reading assignment submissions:

- Copying any segment of code or essay from online sources without citation
- Submitting anything that you did not write yourself or adapt from cited online sources
- Doing anything for the purpose of aiding or assisting another student to commit plagiarism

Students suspected of plagiarism will be referred to the department for formal discipline for breaches of the Code of Behaviour on Academic Matters. There is zero tolerance for plagiarism in this course. For more information on the penalties and offences under the U of T Code of Behaviour on Academic Matters, please refer to

<https://academicintegrity.utoronto.ca/key-consequences/>

Video Privacy and Restrictions on the Course Materials:

Course materials belong to your instructor, the University, the publisher, and/or other sources depending on the specific facts of each situation and are protected by copyright. Lectures of this course, including your participation, might be recorded. Do not download, copy, record, reproduce, or share any course or student materials or videos without obtaining the explicit written permission of the course instructor beforehand.

Acknowledgment of Traditional Land

We wish to acknowledge this land on which the University of Toronto operates. For thousands of years, it has been the traditional land of the Huron-Wendat, the Seneca, and most recently, the Mississaugas of the Credit River. Today, this meeting place is still the home to many Indigenous people from across Turtle Island and we are grateful to have the opportunity to work on this land.

Accessibility:

Disability is a natural part of the human experience. We all experience disability if we live long enough. The University of Toronto and your instructor are committed to accessibility. U of T supports accommodations for students with diverse learning needs, which may be associated with mental health conditions, learning disabilities, autism spectrum, ADHD, mobility impairments, functional/fine motor impairments, concussion or head injury, blindness and low vision, chronic health conditions, addictions, deafness and hearing loss, communication disorders and/or temporary disabilities, such as fractures and severe sprains, or recovery from an operation.

If you have a learning need requiring an accommodation the University of Toronto recommends that students register as soon as possible with Accessibility Services at <https://studentlife.utoronto.ca/service/accessibility-services-registration-and-documentation-requirements/>

Phone: 416-978-8060 Email: accessibility.services@utoronto.ca

Inclusivity Statement:

Students and the instruction team of this course represent a diverse community in which mutual respect is a foremost principle. This means that no matter a person's gender, race, religion, sexual identity, language or background they will be included as equals, their ideas

will be listened to and considered and they will be given all opportunity to participate and contribute to the creation of knowledge that is the goal of this course.

All students and faculty at the University of Toronto have a right to learn, work and create in a welcoming, respectful, inclusive and safe environment. In this class we are all responsible for our language, action and interactions. Discriminatory comments or actions of any kind will not be permitted. This includes but is not limited to acts of racism, sexism, homophobia, transphobia, and ableism. As a class we will work together to create an inclusive learning environment and support each other's learning.

If you experience or witness any form of discrimination, please reach out to the Engineering Equity Diversity & Inclusion Action Group [online](#), an [academic advisor](#), a [U of T Equity Office](#), or any U of T Engineering faculty or staff member that you feel comfortable approaching.

Mental Health:

As a university student, you may experience a range of health and/or mental health challenges that could result in significant barriers to achieving your personal and academic goals. Please note, the University of Toronto and the Faculty of Applied Science & Engineering offer a wide range of free and confidential services that could assist you during these times.

As a U of T Engineering student, you have a Departmental [Graduate Administrator](#) who can support you by advising on personal matters that impact your academics. Other resources that you may find helpful are listed on the [U of T Engineering Mental Health & Wellness webpage](#), and a small selection are also included here:

- [Accessibility Services](#) & the [On-Location Advisor](#)
- [Graduate Engineering Council of Students' Mental Wellness Commission](#)
- [Health & Wellness](#) and the [On-Location Health & Wellness Engineering Counsellor](#)
- [Inclusion & Transition Advisor](#)
- [U of T Engineering's Learning Strategist](#) and [Academic Success](#)
- [U of T Engineering's Mental Health Programs Officer](#)
- [My Student Support Program \(MySSP\)](#)
- [Registrar's Office](#)
- [SKULE Mental Wellness](#)
- [Scholarships & Financial Aid Office & Advisor](#)

If you find yourself feeling distressed and in need of more immediate support resources, consider reaching out to the counsellors at [My Student Support Program \(MySSP\)](#) or visiting U of T Engineering's [Urgent Support – Talk to Someone Right Now webpage](#).