

ECO250Y0: Big Data Tools and Applied Machine Learning for Economists

Instructor: Prof. Khazra

Email: nazanin.khazra@utoronto.ca

Important: please check your Quercus daily and make sure notifications is on. Read the syllabus thoroughly, mark your calendars with important dates, and familiarize yourself with course rules regularly.

1 Course Description and Learning Outcomes

The first half of this course explores unstructured data sources such as text files, webpages, social media posts, satellite imagery, and how economists harness these types of data.

The second half of the course gives an overview of different concepts, techniques, and algorithms in machine learning and their applications in economics. We begin with topics such as classification, linear and non-linear regressions and end with more recent topics such as boosting, support vector machines, and Neural networks as time allows.

This course will give students the basic knowledge behind these machine learning methods and the ability to utilize them in an economic setting. Students will be led and mentored to develop and solve an economic problem with machine learning methods introduced during the course.

By the end of this course, students will be able to:

- Learn to code with Python at intermediate or fluent levels
- Learn to search effectively and debug their code
- Learn coding skills most useful to economists such as GIS mapping, web scraping, and machine learning.
- Understand the process of doing applied Economic research
- Apply their coding knowledge to a real world dataset
- Formulate a research question
- Create a full academic paper from data cleaning to visualization and results using student specific real world datasets

1.1 Required Text and Software

All textbooks and learning materials are available online for free. I use a different source for each section. Here are some useful references that we will selectively use in our course.

- Provided lecture notes and Jupyter notebooks
- Pro Git (Scott Chacon, Ben Straub, 2nd edition) <https://git-scm.com/book/en/v2>
- Introduction to Statistical Learning with Applications in R (James, Witten, Hastie, Tibshirani, 2013) <http://faculty.marshall.usc.edu/gareth-james/ISL/>
- Videos, slides and other material posted on Quercus
- We will mainly use Python and Jupyter notebook.

1.2 Prerequisites

The professor **cannot** change or waive the prerequisites. Please contact the Econ department undergraduate administrative staff if you have any questions.

Prerequisite: ECO100Y1(67%)/(ECO101H1(63%);
ECO102H1(63%))/ECO105Y1(80%)/ECO100Y5(67%)/(ECO101H5(63%);
ECO102H5(63%))/(MGEA02H3 (67%); MGEA06H3 (67%));
MAT133Y1(63%)/(MAT135H1(60%);
MAT136H1(60%))/ MAT137Y1(55%)/MAT157Y1(55%); CSC108H1/CSC148H1
Exclusion: ESC190H

1.3 Online Delivery Requirements

This course will have some online components. The lectures will be a combination of pre-recorded and live synchronous lectures and will be posted on Quercus. Collaboration hours are online.

You need high-speed internet, a PC or laptop.

- Keep a calendar with due dates.
- All times will be posted in local Toronto time, and confusion over time zones will not be considered an appropriate excuse for missing a deadline.
- Take-home assignments are due at 7:00 pm Toronto time on the due date unless otherwise stated.

2 Course Rules

2.1 Email Policy

Before you start writing an email to a member of the course staff:

- Please make sure your question is not already answered in the syllabus or announcements on Quercus

- If this is a coding question:
 - First, try to Google the error that you get (e.g., copy and paste it into Google). Since Python is an open-source program, most of your questions have already been answered on the web.
 - If you could not fix the issue, post it on our discussion platform. Your classmates can learn from your questions. We value active participation (asking and answering questions) on our discussion platform.
 - If you still need more help, attend the collaboration hours and your TAs will answer your questions.
 - At last, if you tried all of the above and still have a question, send an email to the instructor.
- Do not reply to announcement emails
- **Use your UT email address in all your communications**
- Email is mainly for private communications. For content-focused questions, please use the collaboration hours. An alternative way to get answers, show participation, and benefit your classmates is to use our discussion platform.
- **Important:** please write formal emails with proper salutation, body, and closing. There are useful resources in this piece: <https://sociology.utoronto.ca/how-to-write-an-email-when-you-need-help/>.
- If you do not receive a response from me by the end of the next business day, the most likely reason is that one of (a)-(f) above are not satisfied.

2.2 Technical Difficulties Policy

We will not accept missed work due to technical difficulty, deadline confusion, internet, or hardware problems. You can (but try not to) miss one weekly assignment during the semester. Please find the details in section 4; Assignments and Projects. Wisely reserve these options for unforeseen technical difficulties, illness, or other incidents.

3 Course Structure

3.1 Lectures

I will post lectures recordings on Quercus each week. **Do not share any of the course material.** *Course videos and materials belong to your instructor, the University, and/or other sources depending on the specific facts of each situation, and are protected by copyright.* In this course, you are permitted to download session videos and materials for your own academic use, but you should not copy, share, or use them for any other purpose without the explicit permission of the instructor.

We may also have synchronous lecture sessions as needed. Dates of the live sessions will be announced on Quercus. The primary purpose of having the live lectures is to complement the posted material and, most importantly, to have in-person interactions while we are online.

I will provide the lab code and explanation for each week's material. You should run these labs each week after or before the lecture and use them to submit your projects.

3.2 Collaboration Hours and Office Hours

We do not have lecture-based tutorials in this class. However, collaboration hours are going to have the same functionality. You should use these collaboration hours to clarify any questions that you may have about the material, and your projects.

The times and the schedule will be announced on Quercus.

4 Assignments and Projects

Category	Item	Weight	Due Dates ^a	Timing
Projects ^b	Term project 1	18%	Fri. Aug. 11	Take-home
	Term project 2	20%	Fri. Aug. 18	Take-home
	Term project 3	20%	Fri. Aug. 25	Take-home
	Final Project	25%	TBD	Take-home
Active Participation	Active Participation	15%		-
CVs	Resume Submission	2%	Mon. Aug. 7 & Mon. Aug. 28	Take-home

^aAll times mentioned in this table are Toronto times. Due time is at 7:00 PM of the due date.

^bIf you have to miss a term project for a medical reason, your **missed term project's weight** will be distributed between your other two term projects. If you miss more than one term project, it will be marked as zero. Read more in the next section.

Special Accommodation

In case you have to miss an assignment or a project due to illness, technical difficulties, etc., you can use the special accommodation described below. You are strongly advised only to use them if necessary.

1. If you have to miss a term project for a medical reason, your **missed project's weight** will be distributed between your other two term projects. If you miss more than one term project, it will be marked as zero and you may consider dropping the course.

Let's work through some of the implications of this policy.

- For the first term project you miss, there is no need to self-report any reason or illness. No documentation needed/accepted. The weight of the missed project will be distributed between your two other term projects.
- If you miss a subsequent project, you may want to consider dropping the course as you will receive a zero for that project. In the case of extraordinary circumstances, contact your college's Registrar's Office. The only possibility of adjusting the marking policy would be the result of our consultation with your college's Registrar.
- You will have to include all your projects and incorporate the comments in your final project. We point out **there is absolutely no benefit to missing a term project**, even if you cannot submit a perfectly polished work.

- Further, missing a term project is risky as you do not know what the future holds. Assume that for whatever reason, you are going to be forced to miss the Final project.

4.1 Resume Submission

You will submit your resume twice. Once at the beginning of the semester, and the second time at the end of the semester. We are going to provide examples and guidelines on how your resume should look like. Your first resume is going to be evaluated based on the quality and format of the file. In the second submission, you should add your added skills from this course and other courses you took this semester. The guidelines will be posted similarly. Believe me; you will need to send them out to job applications soon!

4.2 Participation

You should consistently and actively participate in the course. Your participation in the course is based on your presence and activity during collaboration hours, the discussion platform, and any live or in-person lectures. This course is project-based, and students encounter different issues; you can increase your participation grade by answering your friends' questions and asking your questions on the platform.

We will track your participation on collaboration hours, the discussion board, and in any live or in-person lectures every week. You should go to the collaboration hours **at three days every week** and answer or ask to questions on the discussion platform.

Ways to get a higher participation mark:

- Attend the collaboration hours every day they are offered. Aside from attending, the coordinator will evaluate your collaboration hours performance; whether you asked a good question, whether you helped a friend solve an error, etc. Attendance with no meaningful contribution does not count toward the participation mark.
- Participate on the discussion board at least once a week. Good and thoughtful answers are weighted more than questions.
- We will track your participation during lectures as well. There may be polls, group work and so on. We award participation marks to high quality and consistent participation through the semester.

I reserve the right not to disclose the distribution of the subcategories of the participation mark as they may change according to the nature of the course and, most importantly, based on the online and in-person division of the course, which is not predictable at the moment. You can calculate your participation mark after the final grades are released on ACORN given your marks for your term projects, CV submissions, and the final project. In addition to the mandatory activities, you can also show your participation in other ways by engaging in opportunities that may come up during the semester.

4.3 Term Projects and The Final Project

You will have three term-projects and a final project. In these projects, you use the provided code and data to finish the defined tasks. We will give you detailed instructions

on the steps required to complete each project. We will also provide feedback on your work, which you should then incorporate and perform the changes that we request. Make sure to address the comments you receive for each project, because you need them for your final project. In the final project, we will add some new parts, and we will also go back and check if you have incorporated our comments into your term projects.

Details about the project will be provided closer to the deadline.

4.4 Projects Late Submission Policy

Late Project submissions will be **penalized by day**. There is a **20% penalty** for each calendar day of late submission. For instance, if the project is due at 7:00 PM of a Wednesday, all the late submissions until Thursday 7:00 PM will incur a 20% point late penalty. There is no grace period. No submissions will be accepted five calendar days after the deadline.

4.5 Remarking Policy

Students should make such requests no later than two weeks after it was returned. Such a request entails a remarking of the entire work and not just the requested part. Hence, if a remarking is granted, the student must accept the resulting mark as the new mark, whether it goes up or down or remains the same. Continuing with the remark or the appeal means the student accepts this condition.

5 Ongoing Learning Disability or Accommodation Requirement

If you have an ongoing disability issue or you need accommodation, please register with Accessibility Services (AS) (accessibility.utoronto.ca) at the beginning of the academic year. After AS processes your request, we will coordinate to provide the required accommodations for you. If you need accessibility related extensions, you should ask your advisor to send us the request at least one week in advance of the due date. We will then coordinate to provide the required accommodations for you.

6 Academic Integrity

The University of Toronto is deeply committed to the free and open exchange of ideas, and to the values of independent inquiry. As such, academic integrity is also fundamental to the University's intellectual life. What does it mean to act with academic integrity? U of T supports the International Center for Academic Integrity's definition of academic integrity as acting in all academic matters with honesty, trust, fairness, respect, responsibility, and courage.

Please visit academicintegrity.utoronto.ca for smart strategies and information on academic integrity processes and procedures at the University of Toronto. The website includes a link to decisions of the University Tribunal in student cases involving academic integrity. You can review the Code of Behaviour on Academic Matters in its entirety [here](#).

Common forms of academic misconduct with code references include:

- Possession or use of unauthorized aids (B.I.1.b). Impersonation (B.I.1.c). Plagiarism (B.I.1.d) (plagiarism is a serious instance of academic misconduct, and university policy explicitly stipulates that ignorance of what constitutes plagiarism is not an acceptable defense.). Submission of work for which credit has previously been obtained (B.I.1.e). Submission of work containing purported statement(s) of fact or reference(s) to concocted sources (B.I.1.f). Assisting another student in committing an offence (B.II.1.a).

7 Online Etiquette

- Do not use your personal email for any course-related activity, registration, or communication.
- When sending any communication or participating in discussions, remember that there are real people with feelings on the receiving end. Be kind and treat people the way you would like to be treated.
- Respect the opinion of your classmates. If you respond to or disagree with your classmates' arguments, do it respectfully and acknowledge the valid points of their arguments.

8 Schedule and Weekly Learning Goals

The schedule is tentative and subject to change. We will try to cover as much of the material as time allows. This schedule should be viewed as a road map to the fundamental concepts that students should learn and study before each assignment.

Week 1

- Introduction to Python, Jupyter notebook
- Linking Different Data Sources
- Version Control with Github
- Data Visualization

Week 2

- Mapping with Python and GIS Mapping
- Satellite Data and Geospatial Visualization
- Satellite Data and Economic Research

Week 3

- HTML-based Web Scraping
- API-based Web Scraping
- Working with Text
- Introduction to Machine Learning
- Linear Regression
- Classification

Week 4

- Cross Validation and Bootstrap
- Shrinkage Methods LASSO and Ridge
- Regression Tree
- Random Forest
- Boosting and Other topics (if time allows)

9 The Final Word

“A university is a place where the universality of the human experience manifests itself.”

Albert Einstein

Everyone is welcome in this class. We should all actively try to create an inclusive environment to give everyone equal chances to grow. I can't do this alone and need your help to achieve this mission. Try to connect with as many of your classmates as the online world allows and welcome new ideas, perspectives, and identities.

I look forward to meeting you all!