

Course Description

This course covers techniques and principles for integrating diverse data sources and methods and tools for knowledge discovery in data. Students will explore methodologies for record linkage, handling missing or inconsistent values, and applying data mining algorithms to uncover interesting trends, and anomalies.

Through their assignments and projects, students will gain practical experience in building data integration pipelines and implementing knowledge discovery techniques to address real-world challenges.

Instructor

Dr. Luiza Antonie,
lantonie@uoguelph.ca

Course Website

courselink.uoguelph.ca

Lectures

M/W 2:30-3:50pm
All class activities
will take place in
person.

Land Acknowledgement

The University of Guelph is on the territory of the Anishnaabe. This is treaty land and territory of the the Missisaugas of the Credit. The Between the Lakes Purchase (Treaty No. 3 of 1792) gave this land to the Haudenosaunee. The University occupies part of this treaty land.

These words recognize that this land has been taken from those who were here first, but they cannot remedy the historical injustices committed here. We recognize and acknowledge those acts, with the intention of a fairer future.

This is also the land of the ancient Treaty of the Dish with One Spoon, and in that understanding, we acknowledge, respect, and share the resources around us and the ideas within us. We recognize that the value of our coming together at this place is to work together, and fairly share in the fruits of our labours.

Course Assessment

		Learning Outcomes	Due Date
Assignments (2)	20%	2,5,6	Oct 11, Nov 8
Paper Presentation and Discussion	10%	1,3	Oct 21/28 Week
Project Proposal	10%	1,3,4	Oct 4
Project Presentations	20%	1,2,4	Nov 25 Week
Project Report	40%	1,2,3,5,6	Dec 6
	100%		

Notes on Course Assessment

- All assignments will be posted on the course web page.
- All students will benefit from a 48 hour grace period after the original due date of the assignments. In all but exceptional circumstances, this should allow enough flexibility to accommodate any unforeseen events that could otherwise impact your work. In fairness to the vast majority of students who are responsible in handing in their work on time, a deduction of 10% will be applied on the first minute that the assignment is handed in after the grace period. An additional 20% deduction will be applied every 24 hours thereafter.
- It is academic misconduct to collaborate on assignments which do not explicitly allow for group work. It is also misconduct to represent other student's work as your own. There are penalties for doing this. All parties to misconduct will be reported to the Director of the School of Computer Science as participants in academic misconduct. See Academic Misconduct in Section VIII of the Undergraduate Calendar for the causes and penalties of misconduct.
- Failure to submit assignments correctly will result in a substantial loss of marks.

Deadlines and Submitting Work

All assignments will be posted and will be submitted electronically through Courouselink. **All deadlines fall at the end of the typical working day (5pm Eastern Time).**

Policy on Re-grade Request

Any request to remark an exam, assignment or quiz must be addressed to your instructor in writing within one week and must include the following (i) your name, your student ID, contact information, and signature, (ii) a clear description of where and why you feel that you were graded in error, and (iii) the following statement exactly as it appears here "In requesting a re-grade I understand that my entire exam/assignment will be re-graded (with particular attention paid to the points that I have brought up) and understand that as a result my score on the exam/assignment could fall as well as rise."

Recommended Textbooks

- **Principles of Data Integration**
Doan, AnHai.; Halevy, Alon.; Ives, Zachary G.
2012
- **Data Mining: Concepts and Techniques**
Jiawei Han, Micheline Kamber and Jian Pei
Morgan Kaufmann, , 3rd ed. 2011.
- **Introduction to Data Mining**
P.-N. Tan, M. Steinbach, and V. Kumar
Addison Wesley, 2005.

Course Learning Outcomes

Knowledge and Understanding:

1. The student should be able to describe fundamental data integration and data mining tasks.
2. The student should be able to demonstrate the ability to understand and implement key algorithms for fundamental data integration and data mining tasks.
3. The student should be able to read, understand and communicate research papers on topics related to the course content.

Skills:

4. The student should be able to effectively communicate the concepts covered in the course.
5. The student will develop problem solving skills, show creativity and critical thinking working on the assignments.
6. The student will develop skills to implement and apply the techniques learnt in class to real world applications.

If You Witness or Are Subject to Unacceptable Behaviour

If you are subject to or witness unacceptable behaviour, or have any other concerns, please notify the course instructor as soon as possible.

If you feel that the course instructor cannot or will not provide remedy for the situation, please contact any of these alternate resources:

Associate Director
(Undergraduate)
ugradir@
socs.uoguelph.ca

Director of the School
director@
socs.uoguelph.ca

Associate Dean
(Academic)
cpesada@
uoguelph.ca

Office of Diversity and
Human Rights
dhrinfo@
uoguelph.ca
or extension 53000

Campus Community
Police at extension
52245

Code of Conduct*

Our learning environment must be a friendly, safe and welcoming environment for all, regardless of ethnicity, gender, sexual orientation, age, ability, socioeconomic status, and religion (or lack thereof). As we wish to facilitate and encourage the fullest participation from everyone, this code of conduct outlines the expectations for all participants (including the instructor and other staff).[‡] This policy is aligned with the larger University policy on Non-Academic Misconduct.

Expected Behaviour

Participate in an authentic and active way. In doing so, you contribute to the health and value of this community.

Exercise consideration and respect in your speech and actions.

Attempt working together before conflict.

Refrain from demeaning, discriminatory, or harassing behaviour and speech.

Be mindful of your surroundings and of your fellow participants.

Alert community leaders (e.g., teaching staff) if you notice a dangerous situation, someone in distress, or violations of this Code of Conduct, even if they seem inconsequential.

Citizenship and Participation

Communities mirror the societies in which they exist and positive action is essential to counteract the many forms of inequality and abuses of power that exist in society.

If you see someone who is **making an extra effort** to ensure our community is welcoming, friendly, and encourages all participants to contribute to the fullest extent, **we want to know!**

Unacceptable Behaviour

Unacceptable behaviours include: intimidating, harassing, abusive, discriminatory, derogatory or demeaning speech or actions by any participant in our community, either in person, online, at any related events, or in one-on-one communications carried out in the context of community business.

Harassment includes: harmful or prejudicial verbal or written comments related to race, religion, disability, sex, gender, sexual orientation; inappropriate use of nudity and/or sexual images in public spaces (including computer labs and presentation slides); deliberate intimidation, stalking or following; harassing photography or recording; posting (or threatening to post) other people's personally identifying information (a.k.a.: "doxing"); sustained disruption of talks or other events; inappropriate physical contact, and unwelcome sexual attention.

Consequences of Unacceptable Behaviour

Unacceptable behaviour from any community member, including the course instructor and those members with decision-making authority, will not be tolerated.

Anyone asked to stop unacceptable behaviour is expected to comply immediately.

If a community member engages in unacceptable behaviour, action *will* be taken to ensure that such behaviour ends, beginning with action on the part of the course instructor, and escalating if necessary.

Additional information on University policy regarding harassment, conduct and human rights is available at the University Human Rights Office.

*This section was developed by Professor A. Hamilton-Wright

‡This code of conduct is based upon the citizen code of conduct available via <http://citizencodeofconduct.org/>, and is distributed under a Creative Commons Attribution-ShareAlike license (<http://creativecommons.org/licenses/by-sa/3.0/>)

Anti-plagiarism Tools

Please note that materials submitted for grading could be evaluated using "MOSS" (Measure of Software Similarity) and "TurnItIn" tools, which detect plagiarism, illicit collaboration and reuse of materials, all of which constitute academic misconduct.

Academic Misconduct

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.

In particular, **it is your responsibility** to accurately and clearly indicate the work of *any and all contributing people*, including yourself, in all presented and submitted materials. By handing in any work for this (or any) course, unless you have specifically identified any other authorship, **you are claiming that the sole authorship is your own**. Including work from any other person without directly indicating the source of such work constitutes academic fraud of some type.

The consequences of any type of academic fraud at the University are significant, and all cases that arise are submitted to the Dean's office for judicial review. **Please do not participate in plagiarism or any other type of academic fraud. It wastes valuable time and resources, which steals energy from the class.**

If you have any questions about what academic fraud such as plagiarism entails, or about any other forms of academic misconduct, please ask your course instructor, the library, or any other faculty. We will be very happy to help you understand this important concept.

Disclaimer

Please note that the ongoing COVID-19 pandemic may necessitate a revision of the format of course offerings, changes in classroom protocols, and academic schedules. Any such changes will be announced via CourseLink and/or class email. This includes on-campus scheduling during the semester, mid-terms and final examination schedules. All University-wide decisions will be posted on the COVID-19 website (<https://news.uoguelph.ca/2019-novel-coronavirus-information/>) and circulated by email.

Illness

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).

Safety Protocols

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.

E-mail Communication

As per university regulations, all students are required to check their *uoguelph.ca* e-mail account regularly: e-mail is the official route of communication between the University and its students. When contacting the TAs and instructor by e-mail, please be sure to use your University of Guelph e-mail address. We apologize not to be able to reply to e-mail inquiries sent from outside the university.

Recording of Materials

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

Drop Date

Students will have until the last day of classes to drop courses without academic penalty. The deadline to drop two-semester courses will be the last day of classes in the second semester. This applies to all students (undergraduate, graduate and diploma) except for Doctor of Veterinary Medicine and Associate Diploma in Veterinary Technology (conventional and alternative delivery) students. The regulations and procedures for course registration are available in their respective Academic Calendars.

When You Can Not Meet a Course Requirement

When you find yourself unable to meet an in-course requirement because of illness or compassionate reasons please advise the course instructor (or designated person, such as a teaching assistant) in writing, with your name, student ID, and e-mail contact. The grounds for Academic Consideration are detailed in the Undergraduate and Graduate Calendars.

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 7 days in advance and not later than the 40th Class Day. More information can be found on the [SAS website](#).

Resources

The [Academic Calendars](#) are the source of information about the University of Guelph's procedures, policies, and regulations that apply to undergraduate, graduate, and diploma programs.