COURSE INFORMATION
COMP-4311 Big Data
Winter 2024, Orillia

Calendar Description:
Students will learn how to select data from big data repositories (like Canadian or US Open Data portals) and how to use proper data analytics techniques. Many data analytics techniques will be introduced such as linear and logistics regression models, decision trees, support vector machine (SVM), association rules and social graph mining. Students will also learn statistical techniques to evaluate the performance of the models.

Instructor Information:

Instructor: Dr. Chris Brogly
Office Location: OA, 3rd floor, #13
E-mail: cbrogly@lakeheadu.ca
Office Hours: Location and hours to be posted on myCourseInfo

TA(s): see myCourseInfo

Course Identification:

Course Delivery: Lectures, MyCourseLink/D2L
Room and Time: OA-2017, 11:30-12:50PM MW
Final Date to Withdraw (Drop): Friday, March 8, 2024
Final Day of Classes: Tuesday, April 9, 2024
Winter Study Week: February 19, 2024 – February 23, 2024

Prerequisite(s):
Computer Science 2477 Object Oriented Programming

Evaluation

Final grades are calculated based on the following evaluation:

- 3-4 Assignments, 15%
- Midterm 1, 20%, TBA, in-class

    - Format: Mixed
- Midterm 2, 25%, TBA, in-class
    - Format: Mixed
- Project, 40%
    - Project Report and Deliverables Evaluation: 40%
    - Due: TBD
    - The instructor will provide suitable options for Big Data projects; students may also select and pursue their own topic with instructor approval. Students will design, prototype, and, ideally, implement a software solution that addresses Big Data. The scope of the project will be limited to be suitable for a one-term course. The instructor will assess the overall quality of deliverables based on the selected topic and a assign a grade. Students may work individually or in groups no larger than 3. Team size may be considered as part of the instructor's assessment. No peer evaluation will be completed; one mark will be assigned to all members. A substantial report on the project will be required.

Course Text and Online Resources:

- Mining of Massive Datasets by Jure Leskovec, Anand Rajaraman, Jeff Ullman. (required)
- The Big Data Textbook by Ghislain Fourny. (required)
- Practical Data Science with Hadoop® and Spark: Designing and Building Effective Analytics at Scale by Casey Stella, Doug Eadline, and Ofer Mendelevitch. (required)
- Research Articles and instructor provided documents will be posted on MyCourseLink. Recommended readings per week will be posted by the instructor on MyCourseLink.

Weeks	**Topics to cover** Note: While all listed topics will be covered – the instructor reserves the right to re-arrange some of the weekly topics as the course progresses.
Week 1	Introduction (Ch 1, Fourny, Ch.1 MMDS)
Week 2	Map-Reduce and the New Software Stack (Ch. 2 MMDS)
Week 3	Data Models and Databases (Ch 2, 5, 7 Fourny)
Week 4	Supervised Learning (Ch. 12 MMDS)
Week 5	Clustering/Unsupervised Learning (Ch. 7 MMDS)

Week 6	Document Similarity (Ch. 3 MMDS), Midterm 1
Week 7	Break (Winter Reading Week Feb 19-23)
Week 8	Graph Databases (Ch. 13 Fourny)
Week 9	Natural Language Processing in Big Data Context (instructor slides)
Week 10	Link Analysis and Web Mining (Ch. 5 Part 1 MMDS)
Week 11	Data Streams (Ch. 4 Part 1 MMDS), Midterm 2
Week 12	Selected topics

Assignments and Evaluations:

Students taking this course must understand and agree that:

(1) Unless otherwise allowed by the course instructor, students must complete the assignments in this course without the assistance of anyone else.

(2) Unless otherwise allowed by the course instructor, students must not access any sources or materials (in print, online, or in any other way) to complete any course exam.

Academic Integrity:

Students must further understand and agree that, if they violate either of these two rules, or if they provide any false or misleading information about their completion of course assignments or exams, they may be prosecuted under the Lakehead University Student Code of Conduct – Academic Integrity, which requires students to act ethically and with integrity in academic matters and to demonstrate behaviors that support the University's academic values.

Assignments: There is **Zero-Tolerance** for plagiarism cases. All such cases will be dealt with according to University prescribed rules. All assignments are individual. Students should understand that their assignments must go through a similarity check and if there is similarity detected then **ZERO** mark will be awarded to those students involved in copying.

Late Assignments: Late assignments will automatically receive a **ZERO** however they will be reviewed to provide formative evaluation feedback and must be submitted for course Completion.

Course Policies:

- Behavioral standards to follow: Student Code of Conduct - Academic Integrity
- Attendance and participation in class discussions is highly recommended.
- Students can communicate with the instructor through email.
- The course outline and schedule are not fixed and subject to change based on class flow.
- University's attendance policy is followed.
- There will be 3-4 assignments. There are late penalties for assignments determined by the course instructor. Extensions will be granted only by the course instructor. If you have medical or compassionate grounds for an extension, you should take supporting documentation to the office of the Dean of your faculty, who will contact the instructor.
- Assignments and time will be posted on MyCourseLink/D2L and announced via emails to all registered students.

Copyright:

Students should be aware that all instructional, reference, and administrative materials prepared for this course are protected in their entirety by copyright. Students are expected to comply with this copyright by only accessing and using the course materials for personal educational use related to the course, and that the materials cannot be shared in any way, without the written authorization of the course instructor. If this copyright is infringed in any way, students may be prosecuted under the Lakehead University Student Code of Conduct – Academic Integrity, which requires students to act ethically and with integrity in academic matters and to demonstrate behaviors that support the University's academic values.

Regulations:

It is the responsibility of each student registered at Lakehead University to be familiar with, and comply with all the terms, requirements, regulations, policies and conditions in the Lakehead University Academic Calendar. This includes, but is not limited to, Academic Program Requirements, Academic Schedule of Dates, University and Faculty/School Policies and Regulations and the Fees and Refund Policies and Schedules (Lakehead University Regulations webpage, 2023-24).

Academic Integrity:

A breach of Academic Integrity is a serious offense. The principle of Academic Integrity, particularly of doing one's own work, documenting properly (including use of quotation marks, appropriate paraphrasing and referencing/citation), collaborating appropriately, and avoiding misrepresentation, is a core principle in university study. Students should view the Student Code of Conduct - Academic Integrity for a full description of academic offenses, procedures when Academic Integrity breaches are suspected and sanctions for breaches of Academic Integrity.

Support for Students:

There are many resources available to support students. These include but are not limited to:

- Health and Wellness
- Student Success Centre
- Student Accessibility Centre
- Library
- Lakehead International
- Indigenous Initiatives

Lakehead University is committed to achieving full accessibility for persons with disabilities. Part of this commitment includes arranging academic accommodations for students with disabilities and/or medical conditions to ensure they have an equitable opportunity to participate in all of their academic activities. If you are a student with a disability and think you may need accommodations, you are strongly encouraged to contact Student Accessibility Services (SAS) and register as early as possible. For more information, please contact Student Accessibility Services (SC0003, 343-8047 or sas@lakeheadu.ca