UNIVERSITY OF GUELPH

DATA*6300/CIS*6180 Analysis of Big Data Winter 2023

---

1 INSTRUCTOR

Instructor: Taiwo Omomule, PhD

Office: MACN 535

E-mail: taiwoo@uoguelph.ca

---

2 AIMS & OBJECTIVES

2.1 Calendar Description

This course introduces software tools and data science techniques for analyzing big data. It covers big data principles, the state-of-the-art methodologies for large data management and analysis, and their applications to real-world problems. Modern and traditional machine learning techniques and data mining methods are discussed, and ethical implications of big data analysis are examined.

2.2 Course Description

This course introduces software tools and data science techniques for analyzing big data. It covers big data principles, the state-of-the-art methodologies for large data management and analysis, and their applications to real-world problems. Modern and traditional machine learning techniques, such as deep learning and data mining methods, are discussed in the context of analyzing very large amounts of data. This course will focus on real-world applications, to equip students with the technical and intellectual skills needed to tackle the challenges and opportunities of big data. The course also discusses the ethical issues associated with the big data life cycle in the context of real-world applications.

Course topics include:

- Introduction to Big Data and its related technologies.

- Data querying (SQL).

- Review of Big data analytics principles, techniques, and tools: distributed processing engines. (e.g. Map Reduce and Stack) and distributed file system (e.g. Hadoop Distributed File System).

- Big data analysis techniques: deep learning, machine learning, ensemble learning, random forest, data cleaning and preprocessing methods, finding similar items, association rule mining, and other supervised and unsupervised methods suitable for large datasets.

- Legal, social, and ethical issues of big data collection and analysis.

2.3 Learning Outcomes

Upon successful completion of this course, students will have demonstrated the ability to:

1. Perform data extraction and management for very large datasets, including (non-geo)spatial-temporal data.

2. Process and analyze large datasets with variety of machine learning, ensemble learning, and data mining techniques.

3. Understand algorithmic techniques and modelling frameworks for processing large amounts of data.

4. Understand, at an advanced level, the various data analytics tools that can be applied to large datasets.

5. Understand, at an advanced level, various steps and tools required to turn large datasets into informed insight.

6. Understand ethical issues related to analyzing big data.

2.4 Instructor's Role and Responsibility to Students

The role of the instructor is to teach new concepts, facilitate discussion, provide feedback to students, and supervise students throughout the design and implementation of the final project.

3 TEACHING AND LEARNING ACTIVITIES

3.1 Timetable

Lectures: Monday, 2:30 - 5:20 PM

Hours: See Courselink; office hours may change.

Venue: See Courselink; TBD

Dates: Jan 8 - April 23, 2024

Course Delivery Format: In-Person (sometimes lectures may be taken remotely if required)

3.2 Course Topics and Schedule

Week	Topic
[Week 1]	Course Overview Introduction to Big Data Data and Analytics Solution Development Process
[Weeks 2-3]	Review of Data Mining and Machine Learning Techniques Project – Phase 1: Selection, Problem Statement, and Analysis Plan
[Weeks 4-5]	Exploratory Data Analysis and Visualization Project - Phase 2: Data Preparation and Analysis
[Weeks 6-7]	Analysis Methods for big data collection and management, including basics of temporal data collected from different locations. Project – Phase 2: Data Analysis and Visualization
[Week 8]	Methods for big data processing and modelling Project – Phase 3: Solution Development Methods
[Weeks 9-10]	Methods for big data processing and modelling Project-Phase 3 & 4: Solution Development, Evaluation, Selection, and Deployment
[Weeks 11-12]	Ethics by Design: Incorporating ethics and governance throughout the Data and Analytics process Project- Presentations

4 Learning Resources

4.1 Course Website

Lecture notes, assignment and project requirements, related course materials,

announcements, and grades will be regularly posted to the DATA*6300/CIS*6180 Courselink account. The students are responsible for checking the website regularly.

4.2 Required Resources

Suggested sources include:

Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal. Data Mining: Practical Machine Learning Tools and Techniques. Fourth Edition. Morgan Kaufmann, 2016.

Sebastian Raschka and Vahid Mirjalili. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Second Edition. Packt Publishing, 2017.

Weekly articles will be provided throughout the semester.

---

5 ASSESSMENT

5.1 Dates and Distribution

Assignment	Due Date	Weighting %	Learning Outcome(s) Assessed
Assignment 1	Week 3	10	1
Assignment 2	Week 5	10	2,3
Readings and Discussions	Weeks 1-10	10	1-6
Exam	Week 6	20	1-4
Project-Proposal	Week 5	5	1-5
Project-Final Presentation	Weeks 11-12	10	1-5
Project-Final project deliverables	Week 12	35	1-6

5.2 Assessment Descriptions

Assignments 1 and 2 are designed in such a way to advance students' hands-on experience with big data processing and analysis. Students develop important skills to use and process massive data into well-structured format, extract metrics and features (e.g. machine learning, deep learning projects, NLP text analysis, and others), and extract meaningful patterns and trends. Examples will be taken from non-spatio-temporal data collected from different sources and locations.

The exam will test students' understanding of course topics covered in the first half of the course. For the course project, students are expected to apply big data tools and methods

that they learn to a real-world problem to advance their skills and understanding of these techniques. They are also expected to assess ethical concerns associated with the real-world application that they select for their project and elaborate on those concerns and potential solutions.

5.3   Course Grading Policies

Accommodation of Religious Obligations: If you are unable to meet an in-course requirement due to religious obligations, please email the course instructor within two weeks of the start of the semester to make alternate arrangements. See the graduate calendar for information on regulations and procedures for Academic Accommodation of Religious Obligations: `https://www.uoguelph.ca/registrar/calendars/graduate/current/genreg/sec_d0e2228.html`

Passing grade: In order to pass the course, students must obtain a grade of 65% or higher on the total mark of all assessments.

---

6   UNIVERSITY STATEMENTS

6.1   Email 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.

6.2   When You Cannot 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, id#, and e-mail contact. See the Undergraduate Calendar for information on regulations and procedures for Academic Consideration.

6.3   Drop Date

Courses that are one semester long must be dropped by the end of the last day of classes; two-semester courses must be dropped by the last day of classes in the second semester. The regulations and procedures for Dropping Courses are available in the Undergraduate Calendar.

6.4   Copies of Out-Of-Class Assignments

Keep paper and/or other reliable backup copies of all out-of-class assignments: you may be asked to resubmit work at any time.

6.5   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 students are 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 make a booking at least 14 days in advance and no later than November 1 (fall), March 1 (winter) or July 1 (summer). Similarly, new or changed accommodations for online quizzes, tests and exams must be approved at least a week ahead of time. More information: `www.uoguelph.ca/sas`

6.6   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 to 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 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.

The Academic Misconduct Policy is outlined in the Undergraduate Calendar.

6.7   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 classmate or a guest lecturer. Material recorded with permission is restricted to use for that course unless further permission is granted.

6.8   Resources

The Academic Calendars are the source of information about the procedures, policies and regulations of the University of Guelph, which apply to undergraduate, graduate and diploma programs.