COMP-4311-WA Big Data

Department of Computer Science

Winter 2024

Instructor Information

Instructor: Dr. Garima Bajwa

Office Location: AT5023

Telephone: 807-342-8010

E-mail: garima.bajwa@lakeheadu.ca

Office Hours: T/Th; 11:30 AM - 12:30 PM or by appointment

Course Identification

Course Number: COMP-4311-WA

Course Name: Big Data

Course Location: BB1054

Class Times: T/Th; 10:00 am to 11:30 am

Prerequisites: COMP-2477, Object Oriented Programming

Course Description/Overview

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.

Course Learning Objectives

By the end of this course, students will be able to identify basic concepts, terminology, theories, models and methods in the field of big data.

Specifically:

- Show substantial understanding of the big data problems.

- Understand the concepts and methods; know the scopes and possible limitations of each method.

- Gather sufficient relevant data, conduct data analytics using scientific methods, and make appropriate connections between quantitative analysis and real-world problems.

- Use advanced techniques to conduct thorough analysis, interpret the results correctly, and draw reasonable conclusions with sufficient explanation and elaboration.

Course Resources

*Course Website(s)*

* D2L

*Required Course Text*

* Moore, D., George P. McCabe, and B. Craig. Introduction to the Practice of Statistics, San Francisco, CA: Freeman, 2012.
* Mitchell, T. Machine Learning, McGraw-Hill, New York, NY, 1997.
* P Wes McKinney. Python for Data Analysis, 2nd/3rd Edition, O'Reilly Media,2017/2022.
* Leskovec, J., Rajaraman A., Ullman, J., Mining of Massive Datasets, 3rd Edition, Cambridge Univ Press
* Selected research publications and course materials shared by researchers.

Course Schedule/Outline

Week 1 – January 8 – 14, Introduction to Big Data

Week 2 – January 15 – 21, Distributions, Probability, Statistics

Week 3 – January 22 – 28, Descriptive & Inferential Statistics

Week 4 – January 29 – ANOVA, Linear/Logistic Regression

Week 5 – February 5 – 11, *Midterm Exam*

Week 6 – February 12 – 18, Machine Learning

Week 7 – February 19– 25, *(Winter Study Week)* *No Classes*

Week 8 – February 26 – March 3, Classification, Clustering

Week 9 – March 4 – 10, Clustering, Model Evaluations

Week 10 – March 11 – 17, Big Data Technologies

Week 11 – March 18 – 24, Hadoop, HDFS

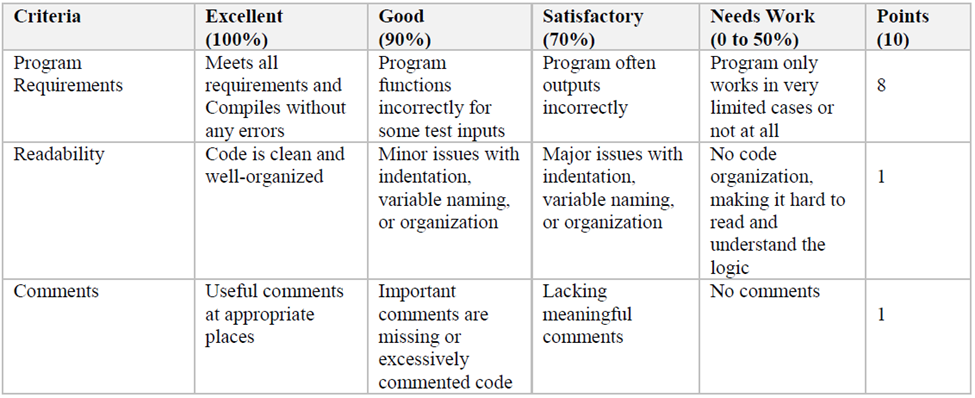
Week 12 – March 25 – 31, MapReduce

Week 13 – April 1 – April 7, Case Study, Review

Assignments and Evaluations

|  |  |  |
| --- | --- | --- |
| Assignment/Test | Date | Value |
| Assignments (programs/theory) + Quizzes | bi-weekly | 30% + 15% |
| Midterm Exam | ~week 6 | 25% |
| Final Exam | Finals week | 30% |

***Example Rubrics for programs***

******

***Late Assignments***

Late submissions will be accepted three days late, with 10% penalty per day and will receive only half credit (50%) after that when turned in. In-class quizzes will not be accepted late.

Submissions will not be accepted after the last class. Any evidence of group participation will be interpreted as academic dishonesty unless it is a group challenge activity/handout.

Course Policies

Submission Policy:

Assignments can be in the form of problem sets or quizzes or programming questions, with a due date posted on D2L.

Email Policy:

Please use only your Lakehead University email for all communication. Include your course title as a prefix in the subject line of your email [Template: COMP-5422-FA: {Subject}]. Send a reminder email if you do not get a response within 48 hours.

Collaboration/Plagiarism Rules

Avoid Plagiarism at all costs. Use citations for ideas that are not yours; Use quoted statements and references; Follow the [Student Code of Conduct - Academic Integrity](https://www.lakeheadu.ca/students/student-life/student-conduct/academic-integrity) - SECTION III: VIOLATIONS OF THIS ACADEMIC INTEGRITY CODE IDENTIFYING OFFENCES (BREACHES OF ACADEMIC INTEGRITY).

If you have any questions, please reach out to me for clarification.

Additionally, posting any course materials (assignments, quizzes, exams, presentations, lecture recording, supplementary resources) to third-party websites without permission is prohibited.

University Policies

Accommodation for Students with Disabilities:

Lakehead University is committed to achieving full accessibility for persons with disabilities. Part of this commitment includes arranging academic accommodations for students with disabilities to ensure they have an equitable opportunity to participate in all their academic activities. If you 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 visit: http://studentaccessibility.lakeheadu.ca