INFS 630 Data Mining (3 credits)

1. Course Description

- Introduction to data mining. Topics include data preprocessing, data warehouse architecture, online analytical processing (OLAP), data lakes, online analytical mining (OLAM), basic concepts and methods of frequent patterns mining, association rules mining, classification analysis, cluster analysis, and text mining.
- Prerequisite: INFS 617 – Information System Design, or basic knowledge of database management systems

2. Learning Outcomes

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

- describe the architecture of data warehouse for supporting online analytical processing
- evaluate different data warehouse and data lake implementations
- describe the mechanisms of major data mining functions
- manually compute data mining results from small sample datasets
- apply data mining software tools to discover hidden patterns from large volume of data
- analyze the results obtained from data mining software tools

3. Textbook and Lecture Notes

- Suggested textbook: Data Mining: Concepts and Techniques, 4th Edition, by Jiawei Han, Jian Pei, and Hanghang Tong, Morgan Kaufmann, 2022. ISBN 978-0128117606. Here is the ebook link: https://mcgill.on.worldcat.org/oclc/1350247579 (You may need to login as a McGill student or on McGill's network/VPN in order to access it.)
- Optional reference: RapidMiner: Data Mining Use Cases and Business Analytics Applications, 2nd Edition, by Markus Hofmann and Ralf Klinkenberg, 2026. ISBN 9781138197008.
- Lecture notes: PowerPoint slides provided by the instructor.

4. Evaluation

1.	In-class exercises:	5%	
2.	Assignments:	15%	
3.	Project:	30%	
4.	Midterm:	15%	
5.	Final exam:	35%	(will be scheduled during the exam period)

If the student's mark on the final exam is greater than his/her mark on the midterm exam, then the midterm mark will be replaced by the final exam mark.

Note: Students are required to bring a scientific calculator with a "log" key to the class in Week 9 (decision tree) and the final exam. The instructor will provide a gentle introduction on the log function.

5. Tentative Timetable

Date	Week	Class Content	Readings in 4th ed.
Aug. 31	1	• Course information • Overview of data warehousing and data mining	Required readings: Chapters 1, 2.1-2.2.1, 2.5.1, 2.5.2
Sept. 7	2	**Labour day (No classes)**	
14	3	• Overview of data warehousing and data mining (cont'd) • Data types	Required readings: Chapters 3.1, 3.2, 3.3.1, 3.4.1, 3.4.2, 3.4.3
21	4	• Multidimensional data model • Online analytical processing (OLAP)	
28	5	• Online analytical processing (OLAP) (cont'd) + Demonstration • Data warehouse architecture • Data warehouse implementation • Online analytical mining (OLAM) ^^^^^ Midterm materials ^^^^^	
Oct. 5	6	**Election day (No classes)**	
12	7	**Study break (No classes)**	
19	8	• Midterm exam (Location: same classroom) • Data lake	(close notes)
26	9	• Frequent patterns mining • Association rules mining (ARM)	Required readings: Chapters 4.1, 4.2.1-4.2.2, 4.3.1-4.3.2 Optional: Chapter 4.2.4
Nov. 2	10	• Classification analysis: decision tree induction, & Bayes classification	Required readings: Chapters 6.1, 6.2 Optional: Chapter 8.3 (bring a scientific calculator)
9	11	• Classification analysis: SVM, lazy learners, and frequent pattern-based classification	Required readings: Chapters 6.6.1-6.6.4, 6.7.1-6.7.3, 7.3.1
16	12	• Classification analysis: ensemble methods • Cluster analysis: k-means, k-medoids & hierarchical clustering	Required readings: Chapters 8.1, 8.2, 8.3.1-8.3.3, 8.4.1, 8.5.1
23	13	• Cluster analysis: DBScan	
30	14	• Project presentations	
Dec. 3	14	• Data mining / machine learning applications or research topics	(make-up class for labour day)
7	15	• Final exam • Location: BIRKS 205 (same classroom)	(bring a scientific calculator; open notes)

6. McGill Policy Statements

- McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures.
- In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.
- Instructor generated course materials (e.g., handouts, notes, summaries, exam questions, etc.) are protected by law and may not be copied or distributed in any form or in any medium without explicit permission of the instructor. Note that infringements of copyright can be subject to follow up by the University under the Code of Student Conduct and Disciplinary Procedures.
- Students with accessibility concern may contact the instructor to arrange a time to discuss the situation. Students may also consider contacting the Student Accessibility & Achievement at 514-398-6009 before discussing with the instructor.
- Mobile computing and communications devices are permitted in class insofar as their use does not disrupt the teaching and learning process.
- End-of-course evaluations are one of the ways that McGill works towards maintaining and improving the quality of courses and the student's learning experience. Students will be notified by e-mail when the evaluations are available on Mercury, the online course evaluation system. Please note that a minimum number of responses must be received for results to be available to students.
- McGill has policies on sustainability, paper use and other initiatives to promote a culture of sustainability at McGill. (See the Office of Sustainability.)
- In the event of extraordinary circumstances beyond the University's control, the content and/or evaluation scheme in this course is subject to change.

7. Student's Responsibilities

- The in-class exercises (weighted 5%) are conducted without prior announcements. Students are expected to attend every class. There are no substitutions for missing in-class exercises. If a student cannot attend class for any reason, they should talk or write to the instructor as soon as possible.
- Some course materials covered in class may not be available on the course website. Students are expected to read the assigned materials and to participate in class discussions.

8. Liaison Librarian

Marcela Isuster
Liaison Librarian for the School of Information Studies
Digital Scholarship Hub & Liaison Librarian to the School of Information Studies and the department of Hispanic Studies
514-398-4729 | marcela.isuster(*at*)mcgill.ca