GLIS 630 Data Mining (3 credits)

1. Course Description

- Introduction to data mining. Topics include data preprocessing, data warehouse architecture, online analytical processing (OLAP), online analytical mining (OLAM), basic concepts and methods of frequent patterns mining, association rules mining, classification analysis, cluster analysis, and text mining.
- Prerequisite: GLIS 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 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

- Required textbook: Data Mining: Concepts and Techniques, 3rd Edition, by Jiawei Han, Micheline Kamber, and Jian Pei, Morgan Kaufmann, 2012. ISBN 978-0-12-381479-1. You can obtain a free copy via McGill Network or McGill VPN: https://www.sciencedirect.com/science/book/9780123814791
- Optional reference: RapidMiner: Data Mining Use Cases and Business Analytics Applications, by Markus Hofmann and Ralf Klinkenberg, 2013. ISBN 9781482205497 and ISBN 9781482205503.
- Lecture notes: Powerpoint slides provided by the instructor.

4. Evaluation

1.	In-class exercises:	10%
2.	Assignments	10%
3.	Project:	30%
4.	Midterm:	15%
5.	Final exam:	35%

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 7 (decision tree) and the final exam. The instructor will provide a gentle introduction on the log function.

5. Tentative Time Table

Date	Week	Class Content	Readings
Jan. 10	1	• Course information • Overview of data warehousing and data mining	Required readings: Chapters 1, 2.1-2.2.1, 3.5
17	2	• Overview of data warehousing and data mining (cont'd)	Required readings: Chapters 4.1, 4.2, 4.3.4, 4.4.1, 4.4.4.
24	3	• Data types • Multidimensional data model • Online analytical processing (OLAP)	Optional: Chapter 4.5.1
31	4	• Online analytical processing (OLAP) (cont'd) + Demonstration • Data warehouse architecture • Data warehouse implementation • Online analytical mining (OLAM) ^^^^^ Midterm materials ^^^^^	
Feb. 7	5	• Frequent patterns mining	Required readings: Chapters 6.1, 6.2.1-6.2.2, 6.3.1-6.3.2, 6.4 Optional: Chapter 6.2.4
14	6	• Midterm exam • Association rules mining (ARM)	
21	7	• Classification analysis: decision tree induction, & Bayes classification	Required readings: Chapters 8.1, 8.2.1-8.2.3, 8.3 (bring a scientific calculator)
28	8	**Study break**	
Mar. 7	9	• Classification analysis: SVM, lazy learners, and frequent pattern-based classification	Required readings: Chapters 8.5.2-8.5.4, 8.6.1-8.6.3, 9.3.1
14	10	• Classification analysis: ensemble methods • Cluster analysis: k-means, k-medoids & hierarchical clustering	Required readings: Chapters 10.1, 10.2, 10.3.1-10.3.2, 10.4.1, 10.6.1
21	11	• Cluster analysis: DBScan • Review session	
28	12	• Project presentations	
Apr. 4	13	• Project presentations	
TBD	TBD	• Final exam	(bring a scientific calculator)

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. See http://www.mcgill.ca/students/srr/honest for more information.
- 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.
- If you have a disability, please contact the instructor to arrange a time to discuss your situation. You may also consider contacting the Office for Students with Disabilities 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. You 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 10%) are conducted in a team of 2-3 students without prior announcements. Students are expected to attend every class. There are no substitutions for missing in-class exercises. If you cannot attend class for any reason, you 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 actively participate in class discussions.

8. Liaison Librarian

David Greene
Liaison Librarian for the School of Information Studies
Humanities and Social Sciences Library
514-398-5925 | david.greene@mcgill.ca