

GLIS 693 Introduction to Data Mining (3 credits)

Disclaimer: This syllabus is provided for informational use only. Content and assignments may change before the start of the course and may differ between course sections. Students of this course should not use this document but instead retrieve the official version from the McGill course management site.

1. Course Description

- Introduction to data mining. Includes 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.
- Co-requisite: GLIS 657 – Database design and development, or basic knowledge of database management systems

2. Learning Outcomes

By the end of the course, the student should be able to:

- describe the architecture of data warehouse for supporting online analytical processing
- evaluate different data warehouse implementations
- describe the mechanisms of the 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 data mining results obtained from the 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
- Lecture notes: Powerpoint slides provided by the instructor.

4. Evaluation

1. (In-class) exercises:	10%
2. Project:	35%
3. Midterm:	20%
4. Final Exam:	35%

Note: Students are required to bring a scientific calculator with a “log” key to class and the final exam. The instructor will provide a gentle introduction on the log function.

5. Tentative Time Table

Week	Class Content	Readings
1	<ul style="list-style-type: none"> Course information Overview of data warehousing and data mining 	Required readings: Chapter 1
2	<ul style="list-style-type: none"> Multidimensional data model Data warehouse architecture Data warehouse implementation 	Required readings: Chapters 4.1, 4.2, 4.3.4, 4.4.1, 4.4.4.
3	<ul style="list-style-type: none"> Online analytical processing (OLAP) Demonstration of OLAP software 	Optional: Chapter 4.5.1
4	<ul style="list-style-type: none"> Online analytical mining (OLAM) 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
5	<ul style="list-style-type: none"> Frequent patterns mining (cont'd) 	
6	<ul style="list-style-type: none"> Midterm (tentative date) Association rules mining (ARM) ARM for business applications 	
7	<ul style="list-style-type: none"> Classification analysis: decision tree induction, & Bayes classification 	Required readings: Chapters 8.5.2-8.5.4, 8.6.1-8.6.2, 8.6.3, 9.3.1 (bring a scientific calculator)
8	<ul style="list-style-type: none"> Classification analysis: SVM, lazy learners, frequent pattern-based classification & ensemble methods Classification analysis for decision making 	
9	Study break	
10	<ul style="list-style-type: none"> 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
11	<ul style="list-style-type: none"> Cluster analysis: DBScan Text mining (if time permits) Review session 	Required readings: Hierarchical document clustering using frequent itemsets (if time permits)
12	<ul style="list-style-type: none"> Project presentations 	
13	<ul style="list-style-type: none"> Project presentations Final report due 	
TBA	<ul style="list-style-type: none"> 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 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.