COMP 4118/6118: Introduction to Data Mining

Instructor: Dr. Xiaofei Zhang

Spring, 2022

E-mail: xiaofei.zhang@memphis.edu
Office Hours: 13:30–14:30 Th (Dunn Hall 318)
TA: TBD

Web: Canvas
Class Hours: 11:20 AM –12:40 T/Th
Class Room: Dunn Hall 225

Course Description

Data mining has emerged as a major frontier field of study in recent years. Aimed at extracting useful and interesting patterns and knowledge from large data repositories such as databases and the Web, the field of data mining integrates techniques from database, statistics and artificial intelligence. This course will provide a broad overview of the field, preparing the students with the ability to conduct research in the field. The following topics will be covered:

- Association
- Clustering
- Classification
- Data Warehouse
- Data Mining over Data Streams
- Web Databases

Textbook

There is no required textbook for this course. Many course materials refer to the following book:

Data Mining: Concepts and Techniques. Jiawei Han, Micheline Kamber and Jian Pei : Morgan Kaufmann Publishers (3rd edition)

Prerequisites

Prerequisites: Algorithms & Data structures

Lecture Schedule

** This is tentative.*

Week	*Topic*	*Highlights*
1	*Overview*	
2	*Association*	*Mining association rules*
3	*FP-tree*	*Frequent pattern mining*
4,5	*Clustering*	*K-means* *Model-based clustering* *Density-based clustering* *Hierarchical clustering*
6	*Outlier*	*Statistical model* *Distance model* *Density model*
7	*Subspace Clustering*	*Dense unit-based method* *Entropy-based* *Transformation-based*
8,9,10,11	*Classification*	*Support Vector Machine* *Neural network* *Recurrent neural network*
12	*Data Warehouse*	

Assessments & Grading

For COMP4118 enrollment:

- *7 assignments: weight 35%*
- *Project: weight 25%*
- *Midterm: weight 15%*
- *Final: weight 25%*

For COMP6118 enrollment:

- *7 assignments: weight 35%*
- *Project: weight 30%*
- *Midterm: weight 15%*
- *Final: weight 20%*

***Note**: the assignment score will be calculated by the top five scored assignments. Assignments and exams may have different questions for COMP4118 and COMP6118 students.*

Grading Scale

We will calculate final letter grades in two different ways; then each student will receive the higher of the two letter grades. One way is a fixed grading scale, with the following cutoffs:

$A \geq 92\%$ $A- \geq 87\%$ $B+ \geq 82\%$ $B \geq 77\%$ $B- \geq 72\%$ $C+ \geq 67\%$ $C \geq 62\%$

The other way is a curve, with the following percentages of students receiving each grade:

$A : 16\%$ $A- : 16\%$ $B+ : 16\%$ $B : 16\%$ $B- : 16\%$ $C+ : 10\%$ $C : 10\%$

However, we will feel free to give an F to any student who clearly did not put effort into the course (or an A+ to any student with truly exceptional performance).

Course Policies

Plagiarism or cheating behavior in any form is unethical and detrimental to proper education and will not be tolerated. All work submitted by a student (projects, programming assignments, lab assignments, quizzes, tests, etc.) is expected to be a student's own work. The plagiarism is incurred when any part of anybody else's work is passed as your own (no proper credit is listed to the sources in your own work) so the reader is led to believe it is therefore your own effort. Students are allowed and encouraged to discuss with each other and look up resources in the literature, but appropriate references must be included for the materials consulted, and appropriate citations made when the material is taken verbatim.

If plagiarism or cheating occurs, the student will receive a failing grade on the assignment and (at the instructor's discretion) a failing grade in the course. The course instructor may also decide to forward the incident to the Office of Student Conduct for further disciplinary action. For further information on U of M code of student conduct and academic discipline procedures, please refer to: http://www.memphis.edu/studentconduct/misconduct.htm