Department of Computer Science
CMPT 463 Data Mining
Course Syllabus Fall 2014

Instructor: Dr. Ankur Agrawal

Email: ankur at manhattan dot edu

Website: http://ankur-agrawal.com

Office Hours: WR 3:00pm – 4:00pm or by appointment in RLC 203C

Class Hours: MWR in RLC 102, 5:30pm – 6:20pm

Overview

This will be an introductory course with focus on the basic concepts of data mining. The course will cover fundamental aspects and techniques of analyzing large, complex datasets. Students will learn methods to recognize patterns and make predictions.

Learning Goals/Outcomes

- Understand the need for data mining and it's use cases
- Understand methods of association analysis
- Understands methods of classification
- Understand methods of clustering
- Understand methods of anomaly detection

Prerequisites

Knowledge of a high level programming language is desirable.

Textbook

Data Mining: Concepts and Techniques, 3rd Edition, Jiawei Han, Micheline Kamber, Jian Pei, ISBN: 978-0-12-381479-1

Tentative List of Topics

- Introduction to data mining
- Data objects and attribute types
- Basic statistical descriptions of data
- Data visualization
- Measuring data similarity and dissimilarity
- Data preprocessing- cleaning, integration, reduction and transformation
- Mining frequent patterns - Apriori algorithm
- Classification methods
- Cluster analysis
- Outlier detection methods
- Additional topics at the instructor's discretion and time-permitting

Method of Evaluation

- Two paper writing and presentation worth 40%
- Assignments, programming and theoretical, worth 30%
- One exam worth 20%
- Class performance worth 10%

Success in Class

- Read the assigned pages in the book as per the class discussion.
- Do as many exercises as possible even if they are not assigned.
- Ask questions about parts of reading or lecture which you do not understand.
- Get help before you are completely lost. I am available to help you via e-mail, in the classroom, or in my office.

Center for Academic Success

Tutoring and support to students is offered in the Learning Center (DLS 206), Leo Learning Center (Leo 117) and the Writing Center (Mig 203).

Attendance Policy

Attendance in every lecture is mandatory. Being in class on time is equally important. Any absence for valid reason will be required to be supported with proper documentation.

Cheating Policy

Cheating will result in zero credit for all students involved. Cheating on an exam will result in an "F" in the course. Programming assignments may not be solved in collaboration, unless specifically stated in the assignment. You may discuss problems with each other. Where does discussion end and cheating start? You may not copy lines of code from anybody or anywhere. You may not use code in your assignments that you did not write. As a general rule, if you don't understand the code and can't explain the code, you can't use the code.

Policy on Students with Disabilities

Students with Disabilities should contact the Specialized Resource Center with their appropriate documentation, to obtain an "Academic Adjustment/Auxiliary Aid" form. When the student presents this completed form to the professor, the professor will then confer with the student on the fulfillment of the adjustments/aids listed on the form.

Academic Integrity Expectation

In accordance with the Manhattan College policy on Academic Integrity, students are expected to do their own work. If they use somebody else's work, then that fact should be documented. Individual work is to be done individually and not copied from others and it is expected that you will perform all exams without consulting others and do your own work on any assignments. Consulting with others on general approaches to take in an assignment is considered acceptable, but copying assignments from others or working the majority of the assignment together is not acceptable. Of course group work is done in a

group.   See   http://manhattan.edu/community-standards-and-student-code-conduct   for   more
information on Manhattan College policy on Academic Integrity.