

## **CSC367: Introduction to Data Mining – Fall 2009**

### **Summary of the course**

Humans are inundated with data in most fields such as finance, health care, and homeland security. Unfortunately, this valuable data, which cost firms millions to collect and collate, are languishing in warehouses and repositories. Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as artificial intelligence, statistical and mathematical techniques.

This course will provide students with the necessary skills at translating the data into knowledge. Data mining is predicted to be one of the most revolutionary developments of the next decade, according to the online technology magazine ZDNET News. In fact, the MIT Technology Review chose data mining as one of ten emerging technologies that will change the world.

This course will illustrate data mining process and how the technology works with sample live applications of data mining. The following topics will be covered:

- Basic concepts, applications and trends in data mining
- Relationship between data mining, data warehouse, and query tools
- Data preparation for the data mining process
- Model building, algorithms and technology:
  - Supervised learning:
    - Association (association rules)
    - Classification and Prediction (decision trees)
  - Unsupervised learning or Self-Organizing (if time permits)
    - Clustering (k-means)
- Evaluation of the data mining model; comparisons of different data mining models
- Visualization using Data Mining (if time permits)
- Data Mining and Society: Future Directions
  - Data mining, national security, privacy and civil liberties.

### **Textbooks and printed resources**

Optional: Data Mining: Concepts and Techniques, by Han and Kamber, Morgan Kaufman Publishers, 2001, ISBN:1-55860-489-8.

Optional: Data Mining: A Tutorial-based Primer, by R. Roiger and M. Geatz, Addison Wesley Publisher 2003, ISBN: 0-201-74128-8.

### **Prerequisites**

IT223 or any other statistics or advanced math class

## Grading

Grading is based on the manner in which you fulfill the objectives of this course. I will grade all your assignments on a percentage basis, which I will then convert to a letter based on the following scale:

Percentage Grade	Letter Grade
95-100	A
90-94	A-
85-89	B+
80-84	B
75-79	B-
70-74	C+
65-69	C
60-64	C-
55-59	D+
50-54	D
0 - 50	F

The weights of each assignment for contributing to the final grade are as follows:

Assignments	Weight in final grade
Homeworks & Programming Assignments	40%
Midterm	30%
Final Project: Proposal, Report, Presentation	30%

## Class Policies

### *Homeworks & Programming Assignments*

There will be 4 homeworks and programming assignments during the quarter. The assignments will be returned at the beginning of the class following the class in which they are due. Late assignments will be accepted up to one lecture later than the assigned due date with a 25% penalty this penalty will be assessed in full to assignments turned in from the end of class on the day that the assignment is due up until the beginning of next lecture. No assignments will be accepted beyond the beginning of class one lecture beyond the due date. Any submitted documents (homeworks, reports, etc) must be typed and submitted through COL website. Extra credit points will be given for additional problems in assignments and midterm, active participation in the lectures and Discussion Forum.

### *Software*

The use of SPSS and SPSS AnswerTree will be taught in class. There will be also two lab sessions scheduled during the second and fifth week of school.

### *Midterm*

There will be a midterm exam given on Thursday, October 15th, that will be worth 30% of the course grade; the midterm is a closed book and notes exam, but students are allowed to bring a calculator.

### *Final Project*

The final project for this class is more extensive analysis task, chosen by you from among the topics we discuss. Final projects will include a presentation to the rest of the class at the end of the quarter, in place of a final exam. As part of your final project, you will also be asked to critique your classmates projects. These critiques will be collected by me, collated, and passed on anonymously to the presenter.

### *Deliverables for the final project*

Proposal -- (October 22nd) One page proposal describing the problem, the proposed approach, and at least three references other than text book or class notes.

Presentation -- (November 12<sup>th</sup>/November 19th) Each project is to be presented using PowerPoint, and the PPT file will have to be submitted to be published on course web site.

Report-- (November 19th) The electronic copy should be in a zip file consisting of all program source code and report itself.

### *Email Communication*

Regarding email communication: I will reply to email messages within one business day after the day I receive and read them; therefore questions that are only received by me on assignments due date (or late the night before) are not guaranteed replies before the assignment is due. Please plan accordingly and begin the assignments early enough to ask questions and receive answers. Also, please do not email me your programs unless I specifically request it.

### *Attendance*

Students are responsible for all material covered, assignments delivered or received, and announcements made in class sessions that you miss.

---

## **School policies:**

---

### *Online Instructor Evaluation*

Evaluations are a way for students to provide valuable feedback regarding their instructor and the course. Detailed feedback will enable the instructor to continuously tailor teaching methods and course content to meet the learning goals of the course and the academic needs of the students. They are a requirement of the course and are key to continue to provide you with the highest quality of teaching. The evaluations are anonymous; the instructor and administration do not track who entered what responses. A program is used to check if the student completed the evaluations, but the evaluation is completely separate from the student's identity. Since 100% participation is our goal, students are sent periodic reminders over two weeks. Students do not receive reminders once they complete the evaluation. Students complete the evaluation online at <https://mycti.cti.depaul.edu/mycti>

### *Email*

Email is the primary means of communication between faculty and students enrolled in this course outside of class time. Students should be sure their email listed under "demographic information" at <http://campusconnect.depaul.edu> is correct.

### *Academic Integrity Policy*

This course will be subject to the academic integrity policy passed by faculty. More information can be found at <http://academicintegrity.depaul.edu/>

### *Plagiarism*

The university and school policy on plagiarism can be summarized as follows: Students in this course should be aware of the strong sanctions that can be imposed against someone guilty of plagiarism. If proven, a charge of plagiarism could result in an automatic F in the course and possible expulsion. The strongest of sanctions will be imposed on anyone who submits as his/her own work any assignment which has been prepared by someone else. If you have any questions or doubts about what plagiarism entails or how to properly acknowledge source materials be sure to consult the instructor.

### *Incomplete*

An incomplete grade is given only for an exceptional reason such as a death in the family, a serious illness, etc. Any such reason must be documented. Any incomplete request must be made at least two weeks before the final, and approved by the Dean of the College of Computing and Digital Media. Any consequences resulting from a poor grade for the course will not be considered as valid reasons for such a request.

### *Resources for Students with Disabilities*

Students who feel they may need an accommodation based on the impact of a disability should contact the instructor privately to discuss their specific needs. All discussions will remain confidential.

To ensure that you receive the most appropriate accommodation based on your needs, contact the instructor as early as possible in the quarter (preferably within the first week of class), and make sure that you have contacted either:

PLuS Program (for LD, AD/HD) at 773-325-4239 in SAC 220  
The Office for Students with Disabilities (for all other disabilities) at 773-325-7290 Student Center 307