Course Information

IS567
Knowledge Discovery Technologies
Fall 2012
Thursday 5:45PM-9:00PM
Loop Campus, CS&TC 224

Instructor Information

Instructor: Daniela Stan Raicu
Office: CDM Center, Room 718
Office Hours: Monday & Thursday 4:00pm-5:30pm
Other times by appointment
Phone: (312) 362-5512
Fax: (312) 362-6116
Email: draicu@cti.depaul.edu
Home page: http://facweb.cs.depaul.edu/~dstan

Course Description

The greatest challenge facing data warehousing professionals is extracting valuable information from the masses of data in the warehouse. One of the most significant and powerful technologies to address this concern is data mining.

Data mining uses statistical analysis and modeling techniques to uncover patterns and relationships hidden in large databases—patterns that ordinary methods might miss. Data mining is only one step in the knowledge discovery process. Other steps include identifying the problem to be solved, collecting and preparing the right data, interpreting and deploying models, and monitoring the results. The real key to success, however, is to have a thorough understanding of your data and your business.
This course will illustrate the knowledge discovery process and how the technology works with sample applications of data mining. The course will cover the following topics:

- What data mining is and is not (Chapter 1)
- Relationship between data mining, data warehouse, and query tools (Chapter 4)
- Applications and trends in data mining (Chapter 13)
- Data understanding and preparation for the data mining process (Chapters 2 & 3)
- Model building, algorithms and technology:
  - Supervised learning:
    - § Classification and Prediction (Chapters 8& 9)
  - Unsupervised learning or Self-Organizing
    - § Clustering (Chapters 10 &11)
- Data Mining for complex data objects (Chapter 13) –if time permits

Course Objectives

After completing the course, the students will be able to:

- identify basic concepts, terminology, models and methods in the field of data mining
- develop and evaluate different data mining algorithms and summarize the results,
- recommend designs of knowledge discovery systems for specific problems

Required Textbook

Data Mining: Concepts and Techniques, by Han and Kamber, Morgan Kaufman Publishers, Third Edition
Textbook webpage: http://www.cs.uiuc.edu/~hanj/bk3/

Prerequisites: IT223: Data Analysis

Grading

The 3 homework/programming assignments will be worth 30% of the course grade. There will be a midterm exam given on Thursday, October 18th, that will be worth 35% of the course grade; the midterm is closed book and notes. In place of a final exam, there is a final project worth 35% of the final grade: the project proposal will be 5%, the report will be 20%, and the presentation will be worth 10%.The proposal is due on Thursday, October 11th, the final presentations are due on November 8th, and the final reports will be due on Thursday, November 15th.

The summary of the weights of each assignment contributing to the final grade is as follows:

Assignment	Weight in final grade
Homeworks & Programming Assignments	30%
Midterm	35%
Final Project	35%

The final grade will be assigned according to the following scale:

Percentage Grade	Letter Grade	Manner of fulfillment
95-100	A	Excellent
90-94	A-	
85-89	B+	Very Good
80-84	B	
75-79	B-	
70-74	C+	Satisfactory
65-69	C	
60-64	C-	Poor
55-59	D+	
50-54	D	
0 – 50	F	

Homework/Programming Assignments, Midterm, and Final Project Policies

Homework/programming assignments
There will be 3 weekly assignments, which are due at the beginning of class one week or two after they are assigned. 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: https://col.cti.depaul.edu

Extra credit points will be given for additional problems in assignments and midterm, active participation in the lectures and Discussion Forum.

Midterm
There will be a midterm exam given on Thursday, October 18th, that will be worth 35% of the course grade; the midterm is a closed book and notes exam, but students are allowed to bring a calculator.

Final Project
The purpose of the final project is to demonstrate your ability to apply the knowledge and the techniques learned during this course. 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.

Whenever it is possible, it is recommended that the DL students attend the final presentations to participate in the live discussions of the final projects and to complete critiques of the other projects. However, appropriate accommodations through SKYPE or WIMBA will be arranged for the DL students not being able to give the presentations in class; the DL students will still have to submit their critiques on the other projects.

Deliverables for the final project:

• Proposal (October 11th): one page proposal describing the problem, the proposed approach, and at least three references other than text book or class notes.

• Presentation (November 8th): 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 15th): The electronic copy should be in a zip file consisting all program source code and report itself.

Software

SPSS will be taught in class and all the assignments and final projects can be implemented using SPSS, but students may use any data mining tool of their choosing when completing class assignments and final projects.

Attendance

It is expected that you will attend every class; it is the single most important action you can take in mastering the course objectives. You are responsible for all material covered, assignments delivered or received, and announcements made in class sessions that you miss. For distance learning students, this means viewing the classes in a timely manner, participate in the discussion forum and/or Wimba, and being sure to email or call in any questions that you have.

Changes to Syllabus

This syllabus is subject to change as necessary to better meet the needs of the students. Significant changes are unlikely, and will be thoroughly addressed in class. Minor changes, especially to the weekly agenda, are possible at any time. You will be informed of all such changes.

Class Cancellation

Unless DePaul University closes because of weather, we will have class.

School policies:

Online Instructor Evaluation

Course and instructor evaluations are critical for maintaining and improving course quality. To make evaluations as meaningful as possible, we need 100% student participation. Therefore, participation in the School's web-based academic administration initiative during the eighth and ninth week of this course is a requirement of this course. Failure to participate in this process will result in a grade of incomplete for the course. This incomplete will be automatically removed within seven weeks after the end of the course and replaced by the grade you would have received if you had fulfilled this requirement.

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

I expect that you have read and understood DePaul's policy on Academic Integrity: http://academicintegrity.depaul.edu/ It is part of this syllabus; follow it.

Plagiarism

The university and school policy on plagiarism can be summarized as follows: Students in this course, as well as all other courses in which independent research or writing play a vital part in the course requirements, 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 a report, examination paper, computer file, lab report, or other 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 School of Computer Science, Telecommunications and Information Systems. Any consequences resulting from a poor grade for the course will not be considered as valid reasons for such a request. Students must formally request an incomplete by filling out a Request for Incomplete Grade form, available at the CDM main office, and submitting it to me.

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