Summary of the course

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 live 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 2)
- Applications and trends in data mining (Chapter 10)
- Data preparation for the data mining process (Chapter 3)
- Model building, algorithms and technology:
  - Supervised learning:
    - Association (Chapter 6)
    - Classification and Prediction (Chapter 7)
  - Unsupervised learning or Self-Organizing
    - Clustering (Chapter 8)
- Data Mining for complex data objects (Chapter 9) –if time permits

Textbooks and printed resources


Prerequisites

CSC 449 and (CSC 212 or CSC 224).

Grading

The 4 homework/programming assignments will be worth 30% of the course grade. There will be a midterm exam given on Monday, May 3rd, that will be worth 30% of the course grade; the midterm is closed book. There will be a quiz that is worth 5% of the final grade. Make-up quiz and exam will not be given. There is no final exam, but there is a final project worth 35% of the final grade: the project proposal will be 5%, the report and demo will be 20%, and the presentation will be worth 10%. Final presentations and demos will be done on June 7th; the final report will also be due on June 7th.
The final grade will be assigned according to the following scale:

- A , 95 -100%;
- A-, 90 - 94%;
- B+, 85 - 89%;
- B , 80 - 84%;
- B-, 75 - 79%;
- C+, 70 - 74%;
- C , 65 - 69%;
- C-, 60 - 64%;
- D+, 55 - 59%;
- D , 50 – 54%;
- F , <50%;

Class Policies

There will be 4 weekly assignments, which are due at the beginning of class one week after they are assigned. Late assignments will be accepted up to one lecture later than the assigned due date with a 15% 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. All assignments must be either typed or written in non-erasable ink. If your assignment is hand-written make sure your handwriting is legible. If I can’t read what you have written, you will receive a zero for the corresponding assignment.

You can submit the assignment in hard copy or through the distance learning web site; grades will be returned through the distance learning web site as well. If submitting a hard copy, always include at the top of the page your last name, class number and section, and assignment number. Please staple all pages together.

Attendance is not mandatory. However, students are responsible for anything that transpires during a class. Lecture notes are available and they cover the material required for this class; you should be aware that the material discussed in class may be covered in more detail than the lecture notes. Also, problems may be discussed in class that are not posted.

My office hours are as follows: Monday: 5:30pm - 6:30pm; Thursday: 4:00pm to 5:30pm and 9:00pm to 9:30pm.

Regarding email communication: I will reply to email messages by one business day after the day I receive and read them; therefore questions that are only received by me on an assignment’s 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.

Students in the distance learning section of the course are expected to take the midterm exam and make the final presentations at the same time as the on-campus section of the course. If there is not sufficient space in the regular classroom to accommodate all students, another location will be provided for distance learning students. Distance learning students who do not reside in the Chicago area must make arrangements to have their exams proctored at a testing site, and information about arranging the exam will be posted to the announcements during the 4th week of the quarter. Distance learning students are subject to the same assignment deadlines as the students in the regular section of the course.

Course Web site

The course home-page will serve as the central information repository for the course. A link to the Web page for this course is found on the CTI Course Online page. To log onto the CTI COL page visit http://dlweb.cti.depaul.edu/. Assignments, lecture notes, study guides for exams, and other course materials will be available only through the COL site. There is also a discussion forum available on the site that you are encouraged to use.

School policies:
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

This course will be subject to the faculty council rules on the academic integrity policy (30k DOC)

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.