

### ***Course Information***

IS567  
Knowledge Discovery Technologies  
Spring 2014  
Tuesday 5:45PM-9:00PM  
Loop Campus, Lewis 01217  
Course Management System: <http://col.cdm.depaul.edu>

### ***Instructor Information***

Instructor: Daniela Stan Raicu  
Office: CDM Center, Room 718  
Office Hours: Tuesday 4:15pm-5:00pm, 9:00pm-9:45pm  
Other times by appointment  
Phone: (312) 362-5512  
Fax: (312) 362-6116  
Email: [draicu@cti.depaul.edu](mailto: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:
  - o Supervised learning:
    - § Classification and Prediction (Chapters 8& 9)
  - o 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:** IT403: Statistics and Data Analysis

**Grading**

The homework/programming assignments will be worth 40% of the course grade, the midterm will be worth 30%, and the final exam/project will be worth 30%. The midterm exam will be given on Tuesday, May 13<sup>th</sup>. The final exam will be a take-home exam in the form of a project; it will be assigned on Tuesday, June 3<sup>rd</sup>, and it will be due on Tuesday, June 10<sup>th</sup>.

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

Assignment	Weight in final grade
Homework & Programming Assignments	40%
Midterm	30%
Final exam/project	30%

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-	Very Good
85-89	B+	
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 4 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.cdm.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 Tuesday, May 13<sup>th</sup>, 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 (no phones are allowed) and a one page (single-sided) of formulas.

**Final Exam/Project**

The final exam will be a take-home exam assigned on Tuesday, June 3<sup>rd</sup>, and due on Tuesday, June 10<sup>th</sup>. The final exam will be designed to demonstrate your ability to apply the knowledge and the data mining

techniques learned during this course to a dataset relevant to the course topics. The final exam will be written in the format of a technical report outlining the problem, summarizing the data, reviewing relevant literature on the outlined problem, describing the methodology, presenting and visualizing the results, and making recommendations.

***For online students:***

Online students must schedule their midterm exam at the COL website (<https://col.cdm.depaul.edu>) during the time frame specified by your instructor (Midterm: May 13<sup>th</sup> – 17<sup>th</sup>). Students living within the Chicago land area are considered local and will be expected to take their exams at a DePaul University campus. Time slots vary by campus and day. They can also take the exam with the other in-class students at the official exam time. Online students living outside the Chicago land area (remote) will have their exams administered by a qualified proctor. You will need to find an acceptable proctor in your area before you register for your exam. Detailed information about online exams is at <http://blogs.cti.depaul.edu/colwiki/Wiki%20Pages/How%20Do%20I%20Take%20My%20Exams.aspx>

***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 choice when completing class assignments and final projects. There will be two hands-on SPSS tutorials, one on April 8<sup>th</sup>, and one on April 29<sup>th</sup>. The tutorials will be recorded and available to both in-class and online students.

***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 being sure to email or call in any questions that you have.

***For online students:***

Recordings of each lecture will be available a few hours after the “live” class, and can be found at the course website <https://col.cdm.depaul.edu>. Online students are expected to watch the lectures every week and to keep up with the course information posted on the course website.

***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.

***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. If a change occurs, it will be thoroughly addressed during class and posted under Announcements in COL.

***Class Cancellation***

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

***Cell Phones/On Call***

If you bring a cell phone to class, it must be off or set to a silent mode. Should you need to answer a call during class, students must leave the room in an undistruptive manner. Out of respect to fellow students and the professor, texting is never allowable in class. If you are required to be on call as part of your job, please advise me at the start of the course.

## ***School policies:***

### ***Online Course Evaluations***

Instructor and course evaluations provide valuable feedback that can improve teaching and learning. The greater the level of participation, the more useful the results. As students, you are in the unique position to view the instructor over time. Your comments about what works and what doesn't can help faculty build on the elements of the course that are strong and improve those that are weak. Isolated comments from students and instructors' peers may also be helpful, but evaluation results based on high response rates may be statistically reliable (believable). As you experience this course and material, think about how your learning is impacted. Your honest opinions about your experience in and commitment to the course and your learning may help improve some components of the course for the next group of students. Positive comments also show the department chairs and college deans the commitment of instructors to the university and teaching evaluation results are one component used in annual performance reviews (including salary raises and promotion/tenure). The evaluation of the instructor and course provides you an opportunity to make your voice heard on an important issue – the quality of teaching at DePaul. Don't miss this opportunity to provide feedback!

### ***Academic Integrity and Plagiarism***

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

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.

### ***Withdrawal***

Students who withdraw from the course do so by using the Campus Connection system (<http://campusconnect.depaul.edu>). Withdrawals processed via this system are effective the day on which they are made. Simply ceasing to attend, or notifying the instructor, or nonpayment of tuition, does not constitute an official withdrawal from class and will result in academic as well as financial penalty.

### ***Retroactive Withdrawal***

This policy exists to assist students for whom extenuating circumstances prevented them from meeting the withdrawal deadline. During their college career students may be allowed one medical/personal administrative withdrawal and one college office administrative withdrawal, each for one or more courses in a single term. Repeated requests will not be considered. Submitting an appeal for retroactive withdrawal does not guarantee approval.

College office appeals for CDM students must be submitted online via MyCDM.

The deadlines for submitting appeals are as follows:

Autumn Quarter: Last day of the last final exam of the subsequent winter quarter  
Winter Quarter: Last day of the last final exam of the subsequent spring quarter  
Spring Quarter: Last day of the last final exam of the subsequent autumn quarter  
Summer Terms: Last day of the last final exam of the subsequent autumn quarter

### ***Excused Absence***

In order to petition for an excused absence, students who miss class due to illness or significant personal circumstances should complete the Absence Notification process through the Dean of Students office. The form can be accessed at <http://studentaffairs.depaul.edu/dos/forms.html>. Students must submit supporting documentation alongside the form. The professor reserves the sole right whether to offer an excused absence and/or academic accommodations for an excused absence.

### ***Incomplete***

An incomplete grade is a special, temporary grade that may be assigned by an instructor when unforeseeable circumstances prevent a student from completing course requirements by the end of the term and when otherwise the student had a record of satisfactory progress in the course. CDM policy requires the student to initiate the request for incomplete grade before the end of the term in which the course is taken. Prior to submitting the incomplete request, the student must discuss the circumstances with the instructor. Students may initiate the incomplete request process in MyCDM.

- All incomplete requests must be approved by the instructor of the course and a CDM Associate Dean. Only exceptions cases will receive such approval.
- If approved, students are required to complete all remaining course requirement independently in consultation with the instructor by the deadline indicated on the incomplete request form.
- By default, an incomplete grade will automatically change to a grade of F after two quarters have elapsed (excluding summer) unless another grade is recorded by the instructor.
- An incomplete grade does NOT grant the student permission to attend the same course in a future quarter.

### ***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 the Center for Students with Disabilities (CSD) at:

Student Center, LPC, Suite #370

Phone number: (773)325.1677

Fax: (773)325.3720

TTY: (773)325.7296