

Course Information

CSC367: Introduction to Data Mining

Spring 2017

Tuesday, Thursday 11:50am-1:20pm

Loop Campus, CS&TC 00220

Instructor Information

Instructor: Daniela Stan Raicu

Office: CDM Center, Room 718

Office Hours: Monday & Wednesday, 4:00pm-5:00pm, 9:00-9:30pm
Thursday, 10:00am-11:30am

Other times by appointment

Advising hours: Tuesday, 1:30-4:30pm

Phone: (312) 362-5512

Fax: (312) 362-6116

Email: draicu@cdm.depaul.edu

Home page: <http://facweb.cs.depaul.edu/~dstan>

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: Classification and Prediction (Decision Trees & Bayesian Networks)
 - Unsupervised learning: Clustering (k-means and hierarchical clustering)
- Evaluation of the data mining model; comparisons of different data mining models
- Visualization using Data Mining (if time permits)

There are opportunities for research assistantships for the students having the best final projects.

Course Learning Goals

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
- apply data mining algorithms to large datasets
- recommend designs of knowledge discovery systems for specific problems

Textbooks and printed resources

Required: "Introduction to Data Mining". Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Addison-Wesley, 2005. ISBN-10: 0321321367

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 equivalent statistics course

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	30%

Class Policies

Homework & Programming Assignments

There will be 4 homework and programming assignments during the quarter. Work to be submitted for the course is generally due one or two weeks after it was assigned; late submissions are allowed with a 5%, 10%, and 15% penalty for a one day, two days, and three days, respectively. No late work will be accepted after three days since the assignment was due.

The assignments must be submitted online on the D2L site at <https://D2L.depaul.edu>. Only legible, organized homework which shows your work will be graded. Include your name, section number, date, and homework number on the first page of your assignment. ***It is your responsibility to check that your files are uploaded correctly on D2L; you should always keep a copy of your submission.***

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

Midterm

There will be a midterm exam given on Thursday, May 4th, 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.

For online students:

Online students must schedule their midterm exam on the D2L website during the time frame specified by your instructor (midterm exam: May 4th – May 6th). 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. 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 <https://www.cdm.depaul.edu/onlinelearning/Pages/Exams.aspx>

Paper Review

Throughout the quarter, the students will be also provided with a list of research papers related to the data mining concepts discussed in class. Three of these papers will be discussed in class; each student will have to review these papers and participate in the in-class discussions and on the discussion forum (for online-students).

The paper readings will be selected from well-known conferences in data mining such as ACM Knowledge Discovery in Databases (KDD) and IEEE International Conference on Data Mining (ICDM).

Final Project

The final project for this class is more extensive analysis task, chosen by you from among the topics we discuss. The purpose of the final project is to demonstrate your ability to apply the knowledge and the techniques learned during this course. 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 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 (May 9th): One-page proposal describing the problem, the proposed approach, and at least

three references other than text book or class notes.

Presentation (May 30th/June 2nd): 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 (June 6th): The report will be written in a format of a paper (abstract, introduction, literature review, methodology, results, discussion, conclusions and future work). The literature review for the final report consists of reading and summarizing about 5 to 6 published papers on the review topic. While the internet can serve as a good source of information, the DePaul Library also has extensive holdings, most of them available electronically.

Software

The use of SPSS will be taught in class. There will be also three lab sessions scheduled during the second, fourth, and sixth week of school. 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://d2l.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 D2L.

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