

# DSC341: Introduction to Data Mining

**Session:** Fall  
11:50-1:20, 9/6 - 11/13  
Class: CDM 216  
Office: CDM 837

**Instructor:** Jonathan F. Gemmell  
Email: jonathan.gemmell@gmail.com  
Office Phone: (312) 362-1361  
Office Hours: See website

**Course Description:** 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.

**Prerequisites:** IT223 or any other statistics or advanced math class.

**Course Management System:** DePaul University's Desire2Learn system (d2l.depaul.edu).

**News Widget:** The primary form of communication for this class will be the news widget on the D2L. Please make sure you subscribe to the widget and that DePaul has you correct email.

**Forums:** The class forum is the preferred place to ask questions about the class. If you have questions about the assignments, the exams or lecture notes, please post them there. I read these frequently. All students should subscribe to the forums so that you receive email updates.

**Optional Textbooks:** "Introduction to Data Mining". Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Addison-Wesley, 2005. ISBN-13: 9780321321367

**Software:** The statistical package used in this course is R and RStudio.

<b>Grades:</b>	Assignments:	4 x 150 points
	Final Project – Proposal:	100 points
	Final Project – Report:	300 points

<b>Final Grades:</b>	A:	90% - 100%
	B:	80% - 90%
	C:	70% - 80%
	D:	60% - 70%
	F:	less than 60

- Plusses and minuses are given for the upper and lower 3% in a letter's range. There is no A+.

**Assignments:** There will be 4 assignments during the quarter. Late assignments incur a 1% penalty per hour. Any submitted documents (homework, reports, etc) must be typed and submitted through the D2L website. All submissions must be submitted in pdf or docx format.

**Final Project:** Student will propose a final project for the class during week 5. Examples of final projects include a research paper, implementation of a data mining algorithm, or a data science study. More details will be given during week 4. Final projects are due during week 11. Late submission incur a 1% penalty per hour.

**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 online 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. Online students are expected to watch the lectures every week and to keep up with the course information posted on the course website.

**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.

**Online Teaching 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 three weeks. Students do not receive reminders once they complete the evaluation. Students complete the evaluation online in CampusConnect.

**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 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