

Instructor Information

Instructors: Roselyne Barreto Tchoua

(T is silent)

Office: CDM 702

Office Hours: **online – posted on BLUESTAR
on Thursdays 4:00-5:30 pm**

Other times by appointment

Link: <https://depaul.zoom.us/j/94914280197>

Phone: 312-362-6796

Email: rtchoua@depaul.edu

will respond within 48 hours or 2 business days

include DSC 441 in the subject!



Course Information

DSC441

Fundamentals of Data Science (Formerly IS467)

Winter 2020-2021 (Thu. Jan. 07th 2020 – Wed. Mar. 11th, 2020)

Thu 5:45PM - 9:00PM (ONLINE SYNCHRONOUS)

Zoom link: <https://depaul.zoom.us/j/97992840664>

Passcode: RBT (My initials)

Course Description

Data science is a vast and growing field that focuses on the technology, tools and techniques required to discover patterns and relationships hidden in large databases. With statistics, machine learning and many other computing and applied math disciplines coming together, there is a large toolkit available to practitioners. In this course, we introduce the concepts of data science, and the tools to work with data from the early stages of gaining understanding, through data preparation, applying algorithms, and evaluating and communicating results. Lecture modules cover techniques and algorithms, while tutorials give direct instruction with a powerful toolkit, and homework problems provide the practice needed to hone your skills. Data science takes a lifetime to master, but the core concepts can be put to use in a single quarter.

Topics include data and its storage and exploration, data cleaning and preprocessing, making predictions (SVM, decision trees) and automatically discovering structure (clustering, association rule mining).

Prerequisites: IT 403 (intro statistics) or DSC 423 (regression) or ECO 520 (business analytics)

This course assumes that you have had a basic course in statistics.

Lecture schedule on a separate document on D2L

Prerequisites:

IT 403 (intro statistics) or DSC 423 (regression) or ECO 520 (business analytics)

This course assumes that you have had a basic course in statistics along with an introductory programming course (e.g., intro Python).

Learning Goals

Specific learning goals are provided for each module of the course on the D2L website. Overall, by the end of the quarter, students will be able to:

1. Clean, smooth and normalize data, including by accounting for outliers and missing data
2. Choose among clustering algorithms, explain differences between them, and interpret results
3. Choose among classification algorithms, explain differences between them, and evaluate results
4. Identify specific ethical concerns in data mining

Textbook & Course Material

There is a required textbook for this course. This book is a great resource on data science, with detailed information on the algorithms we are covering and much more, with clear explanations and additional context.

Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques, 3rd edition*. Morgan Kaufmann, 2011. (1st ed., 2000) (2nd ed., 2006)

Textbook webpage: <http://www.cs.uiuc.edu/~hanj/bk3/>

Grading:

The homework assignments will be worth 60% of the course grade and the final exam will be worth 20%. You will have four quizzes, worth 20% of your grade. For each quiz, you will have one week and three attempts; your final grade will be the average of your three scores. The first four assignments will be worth 10% each. The fifth assignment will be worth 20% and it 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. It is the most extensive assignment among the five homework assignments.

Summary of the weights of each assignment contributing to the final grade:

| Assignment | Weight in final grade |
|----------------------|-----------------------|
| Homework Assignments | 60% |
| Quizzes | 20% |
| Final Exam | 20% |

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 | |

*****IMPORTANT***** - Graduate students taking courses under the purview of the School of Computing will be graded using A/B/C/D/F (no option for Pass/D/F).

Homework Assignments and Final Exam Policies

There will be 5 homework assignments during the quarter. Work to be submitted for the course is generally due one week after it was assigned except for the last one (two weeks); 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.

Quizzes

There will be four quizzes during the course of the quarter assigned in weeks 3, 5, 7, and 9. Quizzes will be made available on D2L. You will have three attempts, 60 minutes per attempt and a week to complete the quizzes. **The average of the three attempts will be taken NOT the best of the three attempts.**

Final exam There will be a final take-home exam that will be worth **20%**. The exam will be made available on **WEEK 10**, Wednesday, March 10th 11:59 PM and you have until Saturday March 13th 11:59 PM to complete it. The exam and the dataset will be uploaded onto D2L. It will have theory questions, (some straight definitions and some concepts to explain in your own words) and one SPSS problem. There will be a review on week 9.

Software

SPSS will be taught in class and all the assignments can be implemented using SPSS, but students may use any data mining tool of their choice when completing class assignments. There will be two hands-on SPSS tutorials on January 14th (week 2) and February 11th (week 6) and an SPSS practice final session on March 4th (week 9, if needed). (See useful information document for more info on the software).

Students are free to use any software that they are more familiar with to complete the assignments. Students who use other software should submit code along with their answers. However, the feedback will be given primarily for SPSS.

Attendance (Not directly applicable this winter – join Zoom call!)

It is expected that you will attend every class and remain for the duration; it is the single most important action you can take in mastering the course objectives. Coming 15 minutes late or leaving 15 minutes constitutes an absence for the student. You are responsible for all material covered, assignments delivered or received, and announcements made in class sessions that you miss. For distance learning students, if you can join the zoom call it would be best, but recordings will be posted on D2L following the class session. This means that attendance translates into viewing the material in a timely manner, participating in the discussion forum, and being sure to email or call in any questions that you have.

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.

Please always put "DSC441" in the subject line.

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. If a change occurs, it will be thoroughly addressed during class and posted under Announcements in D2L.

Attitude

A professional and academic attitude is expected throughout this course. Measurable examples of non-academic or unprofessional attitude include but are not limited to: talking to others when the instructor is speaking, mocking another's opinion, cell phones ringing, emailing, texting or using the internet whether on a phone or computer. If any issues arise

a student may be asked to leave the classroom. The professor will work with the Dean of Students Office to navigate such student issues.

Civil Discourse

DePaul University is a community that thrives on open discourse that challenges students, both intellectually and personally, to be Socially Responsible Leaders. It is the expectation that all dialogue in this course is civil and respectful of the dignity of each student. Any instances of disrespect or hostility can jeopardize a student's ability to be successful in the course. The professor will partner with the Dean of Students Office to assist in managing such issues.

Cell Phones/On Call

(This is also not directly applicable this winter, however mute yourself during the zoom call if you are not speaking)

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 (make sure to mute yourself). 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

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. Please see <https://resources.depaul.edu/teaching-commons/teaching/Pages/online-teaching-evaluations.aspx> for additional information.

Mental Health and Academic Assistance

Balancing the hard work of achieving your educational goals with the other demands of life is difficult at the best of times. For many of us, for a variety of reasons, things are all the more difficult now. I want to make sure you feel comfortable, not embarrassed, reaching out to me for support. I will also point out where the University has great resources just a phone call or email away. These have been created and maintained for you, so use them. Sometimes people feel like their situation isn't the worst possible, so they assume they do not need help, but don't let that prevent you from reaching out.

- DePaul University Counseling Services – mental health is as important as physical health, and we have professionals just a phone call away:

<https://offices.depaul.edu/student-affairs/about/departments/Pages/ucs.aspx> (call (773) 325-7779 or 911 for emergency).

- The kind people at the Office of the Dean of Students can help you with a wide range of topics, including figuring out if you should withdraw or apply for an incomplete: <https://offices.depaul.edu/student-affairs/about/departments/Pages/dos.aspx> .
- There are lots of additional, more specific resources listed here with the Office of Student Affairs, including crisis hotlines and sexual assault resources (note *Title IX* refers to a law protecting you from sex discrimination, including harassment and assault): <https://offices.depaul.edu/student-affairs/support-services/counseling/Pages/Crisis-Hotlines.aspx> .
- Finally, the associate Dean Dr. Lucia Dettori has offered to be a resource who can direct you to the right office if and when necessary (ldettori@cdm.depaul.edu).

Academic Integrity and Plagiarism

This course will be subject to the academic integrity policy passed by faculty. More information can be found at <https://resources.depaul.edu/teaching-commons/teaching/academic-integrity/Pages/default.aspx>

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.

All students are expected to abide by the University's Academic Integrity Policy which prohibits cheating and other misconduct in student coursework. **Publicly sharing or posting online any prior or current materials from this course (including exam questions or answers), is considered to be providing unauthorized assistance prohibited by the policy. Both students who share/post and students who access or use such materials are considered to be cheating under the Policy and will be subject to sanctions for violations of Academic Integrity.**

This also applies to sharing and copying someone else's work. I will consider the one who shares as having the same fault as the one who copies.

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. The dropping dates can be found at: <https://academics.depaul.edu/calendar/Pages/default.aspx>

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.

All students are required to manage their class schedules each term in accordance with the deadlines for enrolling and withdrawing as indicated in the University Academic Calendar. Information on enrollment, withdrawal, grading and incompletes can be found at: <http://www.cdm.depaul.edu/Current%20Students/Pages/PoliciesandProcedures.aspx>

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 Grades

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. All incomplete requests must be approved by the instructor of the course and a CDM Associate Dean. Only exceptions cases will receive such approval. Information about the Incomplete Grades policy can be found at <http://www.cdm.depaul.edu/Current%20Students/Pages/Grading-Policies.aspx>

Preferred Name & Gender Pronouns

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the quarter so that I may make appropriate changes to my records. Please also note that students may choose to identify within the University community with a preferred first name that differs from their legal name and may also update their gender. The preferred first name will appear in University related systems and documents except where the use of the legal name is necessitated or required by University business or legal need. For more information and instructions on how to do so, please see the Student Preferred Name and Gender Policy at <http://policies.depaul.edu/policy/policy.aspx?pid=332>

Students with Disabilities

Students seeking disability-related accommodations are required to register with DePaul's Center for Students with Disabilities (CSD) enabling them to access accommodations and support services to assist with their success. There are two office locations:

- Loop Campus – Lewis Center #1420 – (312) 362-8002
- Lincoln Park Campus – Student Center #370 – (773) 325-1677

Students who register with the Center for Students with Disabilities are also invited to contact Dr. Gregory Moorhead, Director of the Center, privately to discuss how he may assist in facilitating the accommodations to be used in a course. This is best done early in the term. The conversation will remain confidential to the extent possible.

Please see <https://offices.depaul.edu/student-affairs/about/departments/Pages/csd.aspx> for Services and Contact Information.