

Instructor Information

Instructors: Roselyne Barreto Tchoua

(T is silent)

Office: CDM 714

Office Hours: **online – reserve slots by email
on Mondays 4:00-5:30 pm** (in person or online)

Other times by appointment (online)

Link: <https://depaul.zoom.us/j/96600255743>

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 **In Person**

I will post the material on Mondays by 5PM.

Fundamentals of Data Science (Formerly IS467)

Spring 2024-2025

Final Exam Date: In Class (In Person) 6/10/2024

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 along with an introductory programming course (e.g., intro Python).

Lecture schedule on a separate document on D2L

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

Books

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. It provides 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/>

The following books are recommended for the course.

- Covers the world of R libraries for data manipulation and more on ggplot: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Hadley Wickham and Garrett Golemund. O'Reilly Media, January 2017. <https://r4ds.had.co.nz/>
- Covers data mining techniques in less detail but includes a business perspective: *Data Science for Business*. Provost and Fawcett. O'Reilly Media, 2013.
- Similar coverage to our book but a bit newer, however, it is connected to Weka, a data mining package we are not using. Still a great resource. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th Edition. Witten, Frank, Hall and Pal. Morgan Kaufmann Publishers, 2017.
- Covers ggplot, the visualization library we are using, with piles of helpful examples. *R Graphics Cookbook: Practical Recipes for Visualizing Data*, 2nd Edition. Winston Chang. O'Reilly Media, November 2018. <https://r-graphics.org/>

Software

This course will be run with R, a powerful and flexible data analysis tool that comes out of the statistics community. There are quite a few tools available, including other languages like Python, and point-and-click software like SPSS. We have chosen to stick to one language because that way we can offer detailed tutorials to build your capabilities and make sure everyone emerges from the course with a consistent skillset. R has what you need for this course and is continually developed by an open-source community, so it adapts as new techniques become available. It includes general programming

capabilities, which make it completely extensible and flexible, plus plenty of built-in tools and convenient libraries for data preparation and modeling. Oh, right, and it's free!

Those of you in the Data Science program may have taken Python in other courses and are wondering why not stick with one language. The reason is that education should not be tied to one language. The skills you learn transcend individual languages and so it is important to learn multiple tools for the same functionalities so you can be flexible in your career.

R Studio is a convenient and free way to use R (the language). You can download it at the given link but note in the installation instructions that if you don't have R itself, you'll need to install that too by following another link. I strongly recommend you update R at this point, because R continues to evolve and if you are using an older version, you might get very confusing error messages as you work through the tutorials in the course.

1. RStudio (get the latest version of R from the link as well; get the *Desktop* version):
<https://rstudio.com/products/rstudio/download/>
2. Upgrading R:
You can find plenty of fuller instructions if you want, but the basic steps are simple. Also, it should ask you to update all your installed libraries when it's done and you should.
 - a. Open R itself (not RStudio)
 - b. Install the package for updating R
`install.packages("installr")`
 - c. Load the package just like any other library
`library(installr)`
 - d. Use the its function for running the update
`updateR()`

Grading:

Work in this course will be evaluated principally through five homework assignments. There will also be a final exam to give a rounded assessment of the course concepts. Finally, there are elements based on keeping up with the class: quizzes, discussions and participation.

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

Assignment	Weight in final grade
Quizzes	10%
Discussion and Participation	10%
Homework	50%
Final Exam	30%

Quizzes

Weekly quizzes will be administered through D2L and will test the material in the lecture modules. These quizzes are open book/notes/friends. Consider the questions yourself because the purpose is to test yourself on the material. This is called 'recall practice' and

will help you remember the material. I will review each quiz the following week, but you can still retake to improve your score. Everyone should try to get full credit for these quizzes.

Discussions and Participation

The discussion assignments are similar to what might be done as in-class examples. They are exercises to help you practice concepts, typically with straightforward examples. Follow the directions posted as the discussion assignment, which will typically include making your own post and then responding to others. Upon completion, you will be able to see a video reviewing the problem. Grading will not be strict as the purpose is participation.

Those discussion assignments are part of the Discussions and Participation grade. The participation component of the grade is based on my estimation that you have participated in the class, including being part of discussion in the forums and asking questions. Additional credit can be granted for strong participation in the forums, e.g., actively helping fellow students .

Homework

Homework assignments will be submitted online on the D2L website. No email submissions will be accepted. Submissions after the due date will come with an automatic 20% penalty. Please ask if you need an extension on assignments. Life interferes and I do my best to accommodate. No homework will be accepted three days after the due date.

The same type of discussion with peers that is allowed on the discussion board is allowed with respect to homework assignments, however, each student must turn in original work, i.e., your own write-up. Please see the *Expectations* section for more detail.

Final Exam

The final will be held during exam week. The specific date is at the top of the syllabus. The exam will be available for several days for online students, including a weekend, giving people with full-time jobs time to take it on the weekend.

Consider this your notice of the date of the test. Do not make plans (e.g., travel) for this period that will affect your ability to take the test. The final exam is required for this course. Makeups will only be given in extreme circumstances and requests for a makeup must be made as soon as possible. Documentation must be supplied of the relevant circumstances.

All exam videos will be reviewed. If cheating is confirmed, the student will be referred to the Policy on Academic Dishonesty in the Student Handbook.

Final Grading Note

I aim to make grading fair and focused on feedback. Email me if you think this has not been the case.

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).

Expectations

Contacting Me

Please get in touch if you have questions or would like to schedule a meeting outside office hours. **Email is the best option.** NOTE: I respond quickly when possible, but my policy is that you should receive a response by the night of the next business day (i.e., an email Tuesday gets a response by Wednesday night).

Additionally, please do not email me questions to which the answer is right here on the syllabus (e.g., when is the final?) or posted in a note on the D2L page. I may not answer such emails, so if you do not receive a response, double-check that the answer is not easy to find. When you email me, **include DSC 441 online** in the subject.

My policy below about not debugging student code is especially true by email. It is rarely productive as I cannot see it run or test a small change for you. Do not send me a snippet of code and ask 'what is wrong?'. You will likely not get any reply. See *Expectations\Coding* for what you should do.

Finally - come to my office hours or call me during office hours. I encourage you to get help if you need it. I will not judge you for needing help. I will help you.

Workload

Data Science is a wide topic, and we have a lot to cover, so the course is a significant amount of work. In order to learn how to work with data and run machine learning algorithms, you must practice, like any other skill. I expect you to put in the time making mistakes and figuring things out, and in return you can expect me not to waste your time with unhelpful exercises. Most of the practice comes from **homework** assignments. They are built to closely follow the timing of the course so that you get practice with each

module's material. At the end of the quarter, we are adding less new technical material and using some time for review. Correspondingly, the final assignment is less about new material and more about practicing with the bigger picture.

There is significant work to be done that is not directly graded. First, I expect everyone to complete the **quizzes** and **discussion assignments** because they will help reinforce the material. They should not require too much time, and they are only graded based on satisfactory completion, so I expect everyone in the course should be able to get full credit for those components of the final grade.

Second, there are weekly **tutorials** to demonstrate how to carry out the practices explained in lecture modules. The module videos will explain the concepts and context, but the tutorials will show you how to apply them. *The tutorials include instructions for just about everything you need to complete the homework assignments.* You may need to supplement the tutorials by referring to documentation or looking up additional techniques, and that is important because you need to learn how to do so. *Doing the tutorials attentively is the best way to ensure success in the course.*

Coding

The prerequisites require some familiarity with coding, so I will not be covering the very basics, e.g., concepts of writing instructions to the computer, variables and loops. Learning R and its data toolkits is well worth the time, even if you are already familiar with another language. In most cases, you can work with tutorial examples and figure out how to get assignments done, but I encourage you to go beyond that and learn more about available methods, packages and techniques.

Time spent practicing these skills will not be wasted. If you find coding difficult, that's all the more reason to dig in. Do not feel embarrassed about not getting it immediately – most people don't. Skills like this can only be learned by practice. I can line up the information for you, but I cannot give you the ability to do it. You have to bang your head against it over and over until things start to come into focus (metaphorically, of course... repeated headbanging is probably not great for your vision). Your perseverance will be rewarded, as these tools are incredibly useful.

Finally, note that as a matter of policy, I do not debug student code. Again, cannot learn coding without practicing and making mistakes (banging your head against the wall, frankly). Work with each other, use the recommended books and provided tutorials. Start by making an established example run and then start adding and changing. Note that I am an expert in data mining and not an expert in R anyway. That said, if you hit a wall, I'm happy to try to help. Come to office hours (or screenshare) and always start by telling me what you have already tried (this is always a good idea when asking anyone a technical question, including a boss or IT support). Treat me as a last resort for debugging because you need to develop your own techniques.

Discussions Boards and Homework Discussion

The discussion boards are for you. I generally stay off them but browse to look for issues from time to time. Specifically, what I must check for and cannot tolerate on the boards is academic integrity violations. You may not post answers or answer code to the forums. It is okay to discuss homework, give suggestions for plan of attack or discuss strategies, but everyone must do and submit their own work. The forums are also used for forming project groups and weekly discussion posts. In all cases I expect you all to be courteous to each other and respectful of others' time and effort. If there are issues, I will have to shut the forums down for the whole class.

Tutorials, Quizzes

Like the Discussion boards, Tutorials and Quizzes are there *for you*. Quizzes help you remember key points of information and Tutorials give you examples to work from for your assignments.

Grades

See the Grades section for details on the proportional components of your final grade and policies, e.g., for late assignments.

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.

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.