# DSC 540 – Advanced Machine Learning
# Depaul University, College of Computing & Digital Media (CDM)

### *Course Information*
DSC540 - Advanced Machine Learning
Fall 2019-2020
Wednesday 5:45PM-9:00PM
Loop Campus, Lewis 1216
Course Management System: http://d2l.depaul.edu

### *Instructor Information*
Instructor: Casey Bennett
Office: Lewis 1216
Office Hours: Wednesday, 5:00pm-5:45pm, 9:00pm-9:45pm
Other times by appointment
Phone: (270) 577-0439
Email: cbenne33@depaul.edu
Home page: http://www.CaseyBennett.com

### *Course Description*
The course is for students with prior background in data mining or machine learning techniques. The course will cover advanced modeling techniques, including ensemble learning, extended linear models and kernel methods (PCA, support vector machines), probabilistic graphical models, bayesian networks, mixture and latent variable models, biologically-inspired computing, feature selection and feature engineering techniques, Markov models, and temporal modeling to find patterns over time. First the theoretical foundations of these techniques will be presented and augmented with in-class examples and homework problems. Second, the state-of-the-art research related to these techniques will be presented and augmented with paper reviews that highlight the practical applications of these advanced machine learning techniques. Applications of the models will be presented in various domains, including social computing, visual computing, and biomedical and health informatics.

### *Course Learning Goals*
At the end of the course, students should be able to:

- understand the basics behind each machine learning method, as well as the respective pros and cons, for solving data science problems
- understand how to implement machine learning models on real world data using tools such as Python, Scikit-learn, and Spark
- understand methods to evaluate the performance of machine learning models
- understand how "information" in real world applications can be formulated as different data structures, and the effects of feature selection and other preprocessing techniques
- select, combine, and apply specific machine learning techniques for certain data types and challenges, and understand, explain, and interpret the obtained results
- identify recent trends and open directions in the field of machine learning

### *Recommended Books*
None required, but good resources for background reading:

- *Data Mining: Practical Machine Learning Tools and Techniques* by Witten, Frank, and Hall, 4th

Edition, ISBN 978-0-12-374756-0
    - This book has a focus of practical applications and the use of the WEKA toolkit.
- *Probabilistic Graphical Models, Principles and Techniques* by Daphne Koller and Nir Friedman, ISBN 978-0-262-01319-2
    - This book has a focus on theoretical foundations of probabilistic graphical models http://mitpress.mit.edu/books/probabilistic-graphical-models
    - eBook available through the DePaul library at http://depaul.worldcat.org/title/data-mining-practical-machine-learning-toolsandtechniques/oclc/706802868&referer=brief_results
- *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* by Hastie, Tibshirani, Friedman
    - This book has a focus on theoretical foundations of data mining; PDF available at https://web.stanford.edu/~hastie/Papers/ESLII.pdf

### *Prerequisites*
DSC424 (formally CSC 424) and (DSC 441 –formerly IS467, ECT 584, CSC 578, or CSC478)

### *Grading*
The homework/programming assignments will be worth 30% of the course grade, paper reviews will be 20%, class participation will be 15%, and the final project will be worth 35%.

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

| Assignment | Weight in final grade |
|---|---|
| Homeworks & Programming Assignments | 30% |
| Paper Review | 20% |
| Final Project | 35% |
| Participation | 15% |

**Final Grades:** A: 90% - 100%
B: 80% - 90%
C: 70% - 80%
D: 60% - 70%
F: less than 60

### *Assignments*
Homework/programming assignments
There will be 4 homework assignments during the quarter. Students should plan to submit completed code, as well as a written summary no longer than a page long answering the assignment questions. Work to be submitted for the course is generally due one or two weeks after it was assigned; late submissions are allowed for 2 days (up to 48 hours after the due date/time), with a penalty of -10% per day. No late work will be accepted after two 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.

Paper Review:
Throughout the quarter, the students will be also provided with a research paper each week related to the theory discussed in class. Each student will have to review these papers and participate in class discussion. Additionally, the student will *pick any two* of these papers, and write a 2-3 page paper review to be submitted which they will represent 20% of the final grade. These papers should be cover three things: 1) what problem was being addressed, 2) what machine learning methods the author used, and 3) what results they found. Papers are due the same class date they were assigned for discussion.

The paper readings will be selected from the recent literature in major journals and conference proceedings in the fields of machine learning and data mining. They include but are not limited to the ACM KDD Conference on Knowledge Discovery and Data mining (KDD), Artificial Intelligence in Medicine (AIME), the IEEE International Conference on Data Mining (ICDM), the SIAM International Conference on Data Mining, the IEEE Transactions on Knowledge and Data Engineering (TKDE), and the ACM Transactions on Knowledge Discovery from Data (TKDD).

Final Project:
The purpose of the final project is to demonstrate students' ability to apply the knowledge and the techniques learned during this course. The final project for this class is more extensive analysis task, chosen by students from among the topics we discuss. 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, 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 project. However, appropriate accommodations will be made 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: One-page proposal describing the problem, data, proposed machine learning approach, and at least three references from the academic literature (e.g. peer-reviewed journal articles, use Google Scholar to search).

Presentation: Each project is to be presented using PowerPoint in a modified Pecha Kucha style, and the PPT file will have to be submitted to be published on course web site.

Report: The report will be written in the 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.

Participation
Students will also be graded based on their participation in discussion, both in class and on Online Discussion Forum. There will be a discussion of the paper reading each week, so think of some thoughts or comments as you read. There will also be a number of in-class group activities. The belief is that students learn better when they engage their own curiosity, rather than just engage in rote memorization. So bring your curiosity to class.

For online DL students, they will be expected to make at least two posts each week to the online discussion forum, while in-class students will be expected to make at least one post. Posts can be questions or comments about papers read that week, class topics, or programming aspects. Students can either start new post or respond/comment/reply on other students' posts. Posts should be made prior to the start of class each week (Chicago time, US central time).

### *Software*

The use of Python and the Scikit-Learn package will be taught in class. There will be also a lab tutorial scheduled during the first week of class to familiarize the students with the basics of Python, Scikit, and Apache Spark. Each week will have in-class coding lessons on each topic explaining the nuts and bolts of how to deploy machine learning models in applied settings. The Python/Scikit/Spark stack is a very common suite in real-world data science practice, and also a very in-demand skill set for employers at the moment.

Students may use any data mining/data science/machine learning tool of their choice (R, matlab) when completing class assignments. However, use of Python is strongly encouraged, as code templates in Python will be provided for homework coding assignments. For those using Python, homework can be submitted as either basic python scripts (.py files) or as Jupyter notebooks (.ipynb files) … how to do so will be taught in class.

### *Course Schedule*

The course schedule will be maintained on the course website on D2L.

### *Attendance*

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, 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 for distance learning (DL) students. Online students are expected to watch the lectures every week and to keep up with the course information posted on the course website. Online students are further encouraged during in-class group activities to pause the recording and participate as if they were in class. Write down responses to poll questions presented and send them as an email to the instructor as part of your participation grade.

Also, it is important that online students carefully read the assignments section above and pay attention to anything specifically pertaining to online students or DL students. Also, online students should make sure they participate in the Discussion Forum each week prior to start of class (Chicago time, US central time), making at least two posts, either starting a new post or a comment/response/reply on others' posts. Since you are not in class, the more participation here the better for your participation grade.

### *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.

### *Attitude*
A professional and academic attitude is expected throughout this course. Measurable examples of nonacademic 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*
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 undisruptive 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.

## Course Policies

### *Changes to Syllabus*
This syllabus is subject to change as necessary during the quarter. If a change occurs, it will be thoroughly addressed during class, posted under Announcements in D2L and sent via email.

### *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.

### *Academic Integrity and Plagiarism*
This course will be subject to the university's academic integrity policy. More information can be found at https://offices.depaul.edu/oaa/faculty-resources/teaching/academic-integrity/Pages/default.aspx.

### *Academic Policies*
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

### *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

### *Students with Disabilities*

DePaul University is committed to ensuring equal access to its educational and extracurricular opportunities for students with disabilities. The Center for Students with Disabilities (CSD) offers reasonable academic accommodations and services to support our students. We also serve as a resource to the many university departments that have a responsibility to accommodate students. Please see https://offices.depaul.edu/student-affairs/about/departments/Pages/csd.aspx for Services and Contact Information.

### *Proctored exams for OL courses (if applicable)*

If you are an online learning student living in the Chicagoland area (within 30 miles of Chicago), you will need to come to one of DePaul's campuses to take an exam. Online learning students outside of the Chicagoland area are required to locate a proctor at a local library, college or university. You will need to take the exam within the window your instructor gives. Students should examine the course syllabus to find exam dates and the instructor's policy on make-up exams. Detailed information on proctored exams for online learning students can be found at
http://www.cdm.depaul.edu/onlinelearning/Pages/Exams.aspx