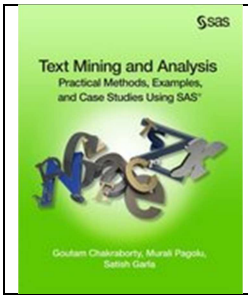
CSC 594 Topics in AI: Text Mining and Analytics, Fall 2015/16

Section numbers: 701 (Loop; class #11805), 710 (Online; class #11806)
Time/location: Section 701 Th 5:45-9:00PM Lewis 1105; Sections 910 Online learning
Instructor: Noriko Tomuro (http://condor.depaul.edu/ntomuro/)
Instructor Email: tomuro@cs.depaul.edu
Office: CDM 648, 312-362-5218
Office hours: M Th 4:15 – 5:15 PM (Loop), Tu 7:00 – 8:00 PM (Online)
Course website: D2L (http://d2l.depaul.edu)

Course Description:
This course will cover the fundamental concepts and practical applications of text mining and analytics. Students will understand of the process pipeline involved, from converting unstructured text to structured data, extracting useful and interesting information such as text topics, trends and user sentiments, then to analyzing the results; and gain practical experience through the use of software tools. Throughout the course, texts from a variety of domains will be used as examples. The course will also introduce some fundamental concepts of Natural Language Processing (NLP) and its techniques and algorithms. Several case studies of real world applications of text mining/analytics will also be covered.

Prerequisite(s): Students should have taken at least one course on Advanced Data Analysis or Data Mining or Machine Learning. Students should also have skills in at least one programming language (e.g. Python, Java, C++) or familiarity with data analytic tools (e.g. SAS, SPSS, MatLab, R).

Textbook:



"Text Mining and Analysis, Practical Methods, Examples and Case Studies Using SAS", G. Chakraborty, M. Pagolu, S. Garla. SAS Institute, 2013.
ISBN-13: 978-1612905518.
https://www.sas.com/store/prodBK_65646_en.html

Important Notes about the book:

- ✓ Although this book is listed as a 'required textbook' at the university bookstore, it is more a recommended reference – thus you do NOT necessarily have to purchase. The book is a practical guide that goes with the software we use in the class (see below). It includes many hands-on, step-by-step examples of how to use the software to do various text mining tasks, along with several interesting case studies. In addition, the book also includes good, concise explanations on the concepts of Natural Language Processing that are relevant in text mining/analytics. For the reasons above, students will probably find the book useful as a course guide.

- ✓ This book is also available for free for DePaul students through Safari Books Online (http://proquestcombo.safaribooksonline.com/book/databases/sas/9781612907871), although only for reading/viewing (i.e., it's not a downloadable ebook).

Software:
In this course, we will use **SAS Text Miner** (http://www.sas.com/en_us/software/analytics/text-miner.html), which is now a component in SAS Enterprise Miner (http://www.sas.com/en_us/software/analytics/enterprise-miner.html) – in particular, the OnDemand (cloud) version provided for free for Academics (http://www.sas.com/en_us/industry/higher-education/on-demand-for-academics.html). Instructions on how to register in the course will be provided by the instructor.

There will be several computer lab sessions in the beginning of the course to let students familiarize themselves with the software. Lab sessions will be recorded so that online students can also keep up with practicing with the software.

Grading:
The grade breakdown will be as follows.

Assignments	50%
Midterm project	20 %
Final project	30 %

All assignments and midterm project must be individual work. Midterm will be a medium-size practical text mining/analytics problem with a focus on the exploration of the software chosen for the course. Final project could be individual or in a small group, and will be a larger project that tackle a real-world text mining/analytics problem. Details of the projects will be announced during the course.

The grading scale will be determined by a curve. The cutoffs will be no higher than the following: 90-100, A; 80-89.99, B; 70-79.99, C; 60-69.99, D; 0-59.99, F. Plusses and minuses will be given at the high/low ends of each grade range (no A+'s or D-'s)

Course Policies

Late Submissions:
Submissions are due at 11:59 pm of the due dates. Late submissions are accepted up to **3 days late**, however will be penalized 10 percent for each day that they are late (including weekend days).

Email Communication:
Any email to the instructor should begin the **subject line with "CSC 594",** so that your message can be easily identified. Failure to do so will delay the response time from the instructor.

Electronic Devices in the Classroom:
Cell phones and other electronic devices must be turned off during lectures and lab sessions. The exception is a laptop, however only for taking notes; other activities that are not related to the current course work, such as emailing, instant messaging, game playing and web browsing, are NOT permitted. If any issues arise, a student may be asked to leave the classroom. The instructor will work with the Dean of Students Office to navigate such student issues.

***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 (D2L).

School 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. Students complete the evaluation online in CampusConnect.

Academic Integrity and Plagiarism:
This course will be subject to the university's academic integrity policy. More information can be found at http://academicintegrity.depaul.edu/. If you have any questions be sure to consult with your professor.

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

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: csd@depaul.edu
Lewis Center 1420, 25 East Jackson Blvd.
Phone number: (312)362-8002
Fax: (312)362-6544
TTY: (773)325.7296