IDSC 4444:
Descriptive and Predictive Analytics
Section 002 - Fall 2018 (Term B)

Classroom: CSOM 1-135
Time: Mon & Wed 11:50AM – 1:30PM
Website: https://canvas.umn.edu/courses/77960

Instructional Staff

Instructor: Yaqiong (Joan) Wang
Contacts:
      Email: wang6285@umn.edu
      Office: CSOM 4-361
      Office Hours: Thursdays 1:30pm – 3:00pm, CSOM 4-361,
            or by appointment

Teaching Assistant(s): Zihong Huang
Contacts:
      Email: huan0707@umn.edu
      Office Hours: By appointment

Course Overview
The past twenty years has witnessed extensive investments in information infrastructure of businesses, which facilitates data collection throughout the enterprise. These data ranges from operations, manufacturing, supply-chain management, customer behavior, etc. With the availability of those data, companies almost in every industry are focusing on analyzing data for competitive advantage. The primary goal of this course is to provide the students with the understanding of the general framework for describing business data and further building and evaluating models to materialize business analytics. It focuses on the fundamental **descriptive** and **predictive** data analytics techniques and methodologies. The course puts significant emphasis on both the understanding of fundamental concepts of data analytics and practical, hands-on experience applying predictive techniques in different settings. Data-analytical thinking and decision making will be advocated in this course through discussion of the value of predictive analytics in various organization contexts and business applications.

Course Goals
Through this course, the students will learn the following:

- General process to perform data analytics.
- Basics of descriptive data analytic techniques including *Data Visualization*, *Data Cleaning* and *Manipulation*.
- Exploratory techniques (unsupervised learning) including *Association Rule Discovery*, *Cluster Analysis* to mine relationships among records of data.
- Predictive models (supervised learning) for *Classification* and *Numeric Prediction* tasks including *Decision Trees*, *K Nearest Neighbors*.
- Prediction performance evaluation-related concepts and topics including *Split Validation*, *Confusion Matrix*, *Lift Curve* and several others.
- Applying predictive modeling techniques in emerging application areas such as *Text Analytics*.
- Basics of *R* programing and how to use R to perform all above data analysis.

Teaching Methods
Classroom

We will alternate each classroom session, i.e., covering lecture material and discussion in every Monday session, and then applying those concepts covered in lectures in hands-on Lab Sessions, completing in-class exercises every Wednesday. **Students are expected to bring their personal laptops to class (either a Mac or PC will be fine)**. During lab sessions, I will demonstrate how to apply the concepts covered in lecture using sample datasets and codes. Students will then work through sample problems in order to prepare for the homework and to gain a better understanding of the subject matter. Lab sessions won't be graded and solutions of in-class exercised would be uploaded at the end of each Wednesday session as your references to homework and exam review.

Software

For the hands on portion of the class, we will be using the R statistical language http://cran.r-project.org/, in tandem with the R-Studio development environment: http://www.rstudio.com/. The R-Studio environment sits on top of the R core, and has the benefit of being much more user friendly. R is a statistical software language that supports an extremely wide array of functions, including a whole host of data mining algorithms. Moreover, R is a free, open source tool, which runs on most major operating systems (including both Windows and Mac OSX). Lastly, R is one of the most popular statistical tools in industry and is also widely used among data scientists, especially statistical learning researchers.

Website

All course related materials, for example, slides, datasets, assignments, quizzes, codes, discussion forums shall be available on the canvas course website, which is available to all registered students. This course does not have a moodle page at the same time. Students are encouraged to get familiar with Canvas. Helpful resources could be found at https://it.umn.edu/canvas-transition-information-student.

To maximize your learning from these experiences, I encourage you to:

- Get started with your homework early, and approach me anytime if you have any questions about the content.
- Approach the tutorials as *learning exercises*, not *button-pushing exercises*.
- Ask questions if you are stuck. The assignments can be complex, so you'll want to get started on the right foot.

Recommended Textbooks / Readings

This course has one recommended textbook, in addition to some supplementary online readings that will be assigned / supplied via Canvas, depending on class progress.

> Provost, Foster, and Tom Fawcett. ***Data Science for Business: What you need to know about data mining and data-analytic thinking***. " O'Reilly Media, Inc.", 2013. http://data-science-for-biz.com/DSB/Home.html
>
> The ebook is available for free to all UMN students through the library here. You are not expected to purchase this book. Specific sections of the book shall be suggested readings for the course per session.

Another book is also highly recommended as an additional reading.

> Shmueli, G., Bruce, P., Yahav, I., Patel, N., Lichtendahl, K., & ProQuest. (2018). ***Data mining for business analytics : Concepts, techniques, and applications in R***. Hoboken, New Jersey: John Wiley & Sons.
>
> This book offers hands-on examples of more advanced descriptive and predictive techniques in R. Digital access through the library is here.

Grading Distribution & Scale

(1) Homework:	50%.
(2) Final Exam:	30%.
(3) Quiz:	20%.
Total:	100%.

Student Deliverables

Homework Assignments: There will be 5 homework assignments, each posted approximately 1 week before the due date specified in the class schedule. These assignments are due before class (on the day your section meets), on the week indicated in the schedule. The detailed information for each assignment can be seen on the Canvas assignments tab [here](). **All assignments should be submitted via Canvas as a single PDF file**. This file should include all subjective answers, calculations, codes, results and plots as requested in the assignment.

- *Late Policy:* If you are late submitting an assignment, it is important for you to complete it and hand it in anyway. Unless waived or reduced by the instructor, late work will be marked 10% lower for each day late up to a maximum of 5 days or 50%. After 5 days beyond the due date (or when sample answers are released), work will no longer be accepted for credit and your grade for the assignment will be zero.

  From a practical standpoint, work submitted after midnight on the due date will be considered one day late, and so on.

- *Format:* Each homework assignment will include conceptual questions, to be answered in written form, and/or hands-on tasks statistical analyses using R. The hands-on tasks will be based on data that I will provide. You will 'mine' the data to get hands-on experience in formulating problems and using the various techniques discussed in class.

Final Exam: The final exam will consist of multiple choice, true/false, and short answer questions, covering the entire breadth of lecture material and lab / homework assignments. The exam will be administered ONLINE, in-class. The exam will be closed book, closed internet and closed notes, but you are free to use a calculator.

Quizzes: There will be 5 short take home quizzes which judge your understanding of the week's concepts and your performance on them will help to identify potential knowledge gaps early on and arrange my before-class reviews. These shall be administered ONLINE using Canvas and will consist of basic multiple choice questions. You shall be able to take the quiz only once, anytime between the Wednesday and Monday sessions. These quizzes will be timed and 10 minutes long. Each quiz shall be based on the preceding week's in-class discussion, slides and exercises and account for 4 points of your grade. The detailed schedule shall be available on Canvas [here]().

Course Guidelines & Academic Policies

Attendance & Collaboration: Students are expected to attend class regularly. As assignment deadlines approach, the demand for help often exceeds supply during office hours and the instructor may not be able to assist all students. To encourage regular attendance, students who regularly attend class will receive preference for help during office hours.

Online Submission: All assignments will use online submission via the course website. Online submission provides a number of advantages to the course staff and students, including date/time stamping and confirmation/verification of your submission. If you are not familiar with the online submission process, please plan to make your first few submissions early so that you have the opportunity to address any issues. The instructor will demonstrate online submission in class before the first assignment is due. For a variety of reasons, you should not email assignments to the instructor or TA unless specifically requested to do so.

Academic Misconduct / Individual Work: Homework should be completed individually. Because the final exam will individually test each student's understanding of the course material, it is in the student's best interest to work through the problems on their own. That being said, you are encouraged to speak with your peers if you are stuck on any of the problems.

For the purposes of this course, academic misconduct is defined as being in possession of an electronic copy of another student's work. This provides a simple mechanism for you to evaluate whether or not you are in compliance with this policy. Submissions may be electronically compared. Work that appears to be an electronic copy, or derived from an electronic copy, of another student's work will be submitted to the respective academic office as evidence of academic misconduct. For the first violation, students will receive no points for the assignment in question. For the second violation in a given semester, the student will receive an F for the current course.

Students should also be careful to protect their own work as when copying occurs it can be difficult to tell who is copying whom. Do not respond to requests by other students for your assignment files. In addition, store your files using secure storage (such as Net Files) and be sure to delete (and empty trash) your files from any public computers you use to complete the assignment.

Accommodations for Students with Disabilities: The University of Minnesota is committed to providing all students equal access to learning opportunities. Disability Services is the campus office that works with students who have disabilities to provide and/or arrange reasonable accommodations. Students registered with Disability Services, who have a letter requesting accommodations, are encouraged to contact the instructor early in the semester. Students who have, or think they may have, a disability (e.g. psychiatric, attentional, learning, vision, hearing, physical, or systemic), are invited to contact Disability Services for a confidential discussion at 612-626-1333 (V/TTY) or at ds@umn.edu. Additional information is available at the DS website (http://ds.umn.edu).

Topics & Schedule

Week	Date	Topic	Assignment Posted	Assignment Due	Quiz Posted
1	10/24 (W)	Course Introduction Working with Data			Getting to know you (Survey)
1	10/29 (M)	Working in R	HW 1		Quiz 1
2	10/31 (W)	Exploratory Analytics 1: Association Rules			
2	11/05 (M)	Lab Session 1: Association Rules	HW 2	HW 1	Quiz 2
3	11/07 (W)	Exploratory Analytics 2: Cluster Analysis			
3	11/12 (M)	Lab Session 2: Cluster Analysis	HW 3	HW 2	Quiz 3
4	11/14 (W)	Predictive Analytics 1: Classification			
4	11/19 (M)	Lab Session 3: Classification	HW 4	HW 3	Quiz 4
5	11/21 (W)	Predictive Analytics 2: Numeric Prediction		ThxGiving	
5	11/26 (M)	Lab Session 4: Numeric Prediction	HW 5		Quiz 5
6	11/28 (W)	Predictive Analytics 3: Text Mining		HW 4	
6	12/03 (M)	Lab Session 5: Text Mining		HW 5	
7	12/05 (W)	Exam Review			
7	12/10 (M)	Final Exam (in-class)			