CpSc 8650 (Section 1), Data Mining  
Spring 2021, Syllabus

General Information:

Class Time:  9:30AM – 10:45AM Tuesday and Thursday.
Location:    Online via Zoom
Instructor:  Dr. James Z. Wang
Office:      125 McAdams Hall
Phone:       864-656-7678
Email:       jzwang@clemson.edu
Office Hours: 9:00AM - 10:00AM Wednesday or by appointment, Online via Zoom.
Final Exam:  Wednesday, April 28, 8:00AM - 10:30AM
Website:     http://www.cs.clemson.edu/~jzwang/21018650/cpsc8650.htm

Prerequisites:

Objective:
Data mining has emerged as one of the most exciting and dynamic fields in computer science, bioinformatics, industrial engineering, etc. The driving force for data mining is the available of massive data that potentially contain valuable bits of hidden knowledge. Such data include consumer data, transaction histories, medical records, biological experiments, Web information, Network information, etc. Commercial enterprises have been quick to recognize the value of data mining; consequently, within the span of a few years, the software market for data mining has expanded to be in excess of tens of billions of dollars. This course is designed to provide graduate students with a broad knowledge in the design and use of data mining algorithms, exposure to data mining research, and hands-on practices in applying these ideas to a real-life situation.

Topics to cover:
This course will cover topics in the emerging area of data mining including (not limiting to) the following:

- fundamental concepts,
- data preprocessing,
- data warehousing,
- information filtering,
- characterization and comparison,
- frequent pattern analysis,
- classification and prediction,
- cluster analysis,
- hypotheses evaluation,
- feature extraction,
- outlier analysis,
- dimensionality reduction, singular value decomposition,
• mining complex data types,
• knowledge discovery in WWW,
• virtualization,
• social networks,
• mobile applications,
• data mining trends and research frontiers.

Special emphasis will be given to multimedia, business, scientific, biological, and medical databases.

**Tentative Course Outline:**
1. Introduction
2. Getting to know your data
3. Data preprocessing
4. Data warehouse and OLAP technology
5. Data cube technologies
6. Characterization and comparison
7. Mining association rules in large databases
8. Classification and prediction
9. Cluster analysis
10. Mining stream, time-series and sequence data
11. Graph mining, social network analysis and multi-relational data mining
12. Mining object, multimedia, text, Web and social network data
13. Data mining trends and research frontiers

**Textbook:**
• Jiawei Han, Micheline Kamber and Jian Pei, Data Mining: Concepts and Techniques, 3rd ed., The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers, July 2011. ISBN 978-0123814791

**Attendance Policy:**
• It is very important to attend the class since the lecture contents may be drawn from various sources.
• Failure to attend the course most likely leads to missing some information that won’t show in lecture notes or textbook.
• Tardy will be punished by penalty in grade.

**Evaluation:**
• **Homework (20%)**: Assignments will be given through the entire semester.
• **Test #1 (20%)**: Cover the contents studied in the first half of the semester.
• **Test #2 (20%)**: Cover the contents studied in the second half of the semester.
• **Project (30%)**: A team project will be assigned to solve a particular data mining problem proposed by the project team.
• **Presentation (10%)**: 15-20 minutes presentation on the project.
• **Penalty Points (-10%)**: Students are required to attend the class. The instructor may penalize those who fail to attend the class without instructor’s approval.
Homework Policy:
- You must finish your homework assignments independently. Any form of cheating will result in 0 point on that assignment for both parties.
- You must submit your homework at the due date. Late submission will not be accepted and will result in 0 point on that assignment.
- Homework will not be returned to you during the semester. However, you can check your homework grading by appointment. Instructor will post the homework solutions regularly to help you comprehend the course content.

Grading:
A (90 - 100), B (80 - 89), C (70 - 79), D (60 - 69), F (0 - 59)

Q&A:
Email questions to jzwang@clemson.edu with subject "CPSC 8650 Q&A".

References:

Academic Integrity:

“As members of the Clemson University community, we have inherited Thomas Green Clemson’s vision of this institution as a ‘high seminary of learning.’ Fundamental to this vision is a mutual commitment to truthfulness, honor, and responsibility, without which we cannot earn the trust and respect of others. Furthermore, we recognize that academic dishonesty detracts from the value of a Clemson degree. Therefore, we shall not tolerate lying, cheating, or stealing in any form.”

“When, in the opinion of a faculty member, there is evidence that a student has committed an act of academic dishonesty, the faculty member shall make a formal written charge of academic dishonesty, including a description of the misconduct, to the Associate Dean for Curriculum in the Office of Undergraduate Studies. At the same time, the faculty member may, but is not required to, inform each involved student privately of the nature of the alleged charge.”

http://www.clemson.edu/academics/academic-integrity/
Accessibility Statement:

Clemson University values the diversity of our student body as a strength and a critical component of our dynamic community. Students with disabilities or temporary injuries/conditions may require accommodations due to barriers in the structure of facilities, course design, technology used for curricular purposes, or other campus resources. Students who experience a barrier to full access to this class should let the professor know, and make an appointment to meet with a staff member in Student Accessibility Services as soon as possible. You can make an appointment by calling 656-6848, by emailing studentaccess@lists.clemson.edu, or by visiting Suite 239 in the Academic Success Center building. Appointments are strongly encouraged - drop-ins will be seen if at all possible, but there could be a significant wait due to scheduled appointments. Students who receive Academic Access Letters are strongly encouraged to request, obtain and present these to their professors as early in the semester as possible so that accommodations can be made in a timely manner. It is the student's responsibility to follow this process each semester. You can access further information here: http://www.clemson.edu/campus-life/campus-services/sds/.

Title IX (Sexual Harassment) Statement:

Clemson University is committed to a policy of equal opportunity for all persons and does not discriminate on the basis of race, color, religion, sex, sexual orientation, gender, pregnancy, national origin, age, disability, veteran's status, genetic information or protected activity in employment, educational programs and activities, admissions and financial aid. This includes a prohibition against sexual harassment and sexual violence as mandated by Title IX of the Education Amendments of 1972. This policy is located at http://www.clemson.edu/campus-life/campus-services/access/title-ix/. Mr. Jerry Knighton is the Clemson University Title IX Coordinator. He also is the Director of Access and Equity. His office is located at 110 Holtzendorff Hall, 656-3184 (voice) or 656-0899 (TDD).