Topic Description:
Advances in connectivity, decreases in the cost of storage, and increases in data generation and data collection have produced an exponential increase in the amount of data that exists. Most of the recorded data is raw, stored in databases in massive data warehouses. Useful information is present in this raw data through patterns—but they must be discovered. Data mining is the extraction and discovery of useful information from data. The goal is that strong identified patterns will generalize and be able to make accurate predictions on future data. Applications of data mining surround us in today’s world: from stock price forecasting, shopping pattern analysis with point-of-sale data (what you buy at the grocery store), prognoses for patients from health data, to personalized recommendations at Amazon.com. The combination of good data with good data mining can be very financially lucrative for both businesses and science.

Meeting Times and Location:
Mondays and Fridays, 9am-9:50am and Wednesday 8am-9:50am in Roddy 256 (Wednesdays we may be in Windows lab)

Office Hours: Mondays and Fridays 11am-noon and 1pm-2pm, and Wednesdays noon-1pm. During office hours I can be found either in one of the labs or in my office (Rm 133, Roddy Hall). My email is stephanie.schwartz@millersville.edu.

Prerequisite: CSCI 362 and MATH 235 or Permission of Instructor


Goals: There are several goals for this course. At the end of this course, the successful student will be able to:

- Understand the fundamentals of data mining, including what kinds of data can be mined, what kinds of patterns can be mined, and what kinds of applications are targeted
- Understand the underlying mathematical and statistical methods used in data mining
- Apply machine learning, pattern recognition, statistics, visualization, algorithm, and database technology in data mining applications
- Evaluate the appropriateness of various technology and algorithms for different domains and problems
- Analyze data in both an exploratory and targeted manner

These goals will be accomplished through the content of the lectures and textbook, as well as hands-on experience. This hands-on experience includes writing programs (both in the lab and in project assignments). There will also be a significant course project in which you identify an analysis topic, discover data, model the data using data mining techniques, analyze the results, and report outcomes. The achievement of the goals will be measured through your performance on approximately 7 lab assignments, the project, and two exams (midterm and final).

Grading:
Midterm: 20%
Final: 20%
Labs and assignments: 30%
Project: 30%

Grading will be on a 100 point scale, with 93%=A, 90%= A-, 87%=B+, 83%= B, etc. You must complete all exams, labs, and assignments in order to pass the course.

Attendance
You are expected to attend class and lab, read the textbook, complete assigned problems, participate in class discussions, and work productively during problem sessions and labs. Lab attendance is REQUIRED. Not attending a lab or leaving before completing all outstanding assignments will result in a grade reduction of 2% per class period missed.

Classroom Civility
Arrive at class promptly, prepared and ready to participate. Set your phone to silent mode and, except in cases of emergency, remain in the classroom for the duration of the meeting. If it is necessary to leave or enter a room once class has begun, do so quietly and with as little disruption as possible. Avoid talking which may be disruptive to your fellow students and professor.

Deadlines
Assignments are due at the beginning of the class period on the assigned due date, unless otherwise specified. No late assignments will be accepted. If your assignment is incomplete, turn it in for possible partial credit.

Exams
There are no make-up exams - if you miss a test, you will receive a zero. Exceptions may be made at my discretion for reasons of illness (as in "on your deathbed") or university excused absences.

Title IX Reporting Obligations
Millersville University and its faculty are committed to assuring a safe and productive educational environment for all students. In order to meet this commitment, comply with Title IX of the Education Amendments of 1972, 20 U.S.C. §1681, et seq., and act in accordance with guidance from the Office for Civil Rights, the University requires faculty members to report to the University’s Title IX Coordinator incidents of sexual violence shared by students. The only exceptions to the faculty member’s reporting obligation are when incidents of sexual violence are communicated by a student during a classroom discussion, in a writing assignment for a class, or as part of a University-approved research project. Faculty members are obligated to report to the person designated in the University Protection of Minors policy incidents of sexual violence or any other abuse of a student who was, or is, a child (a person under 18 years of age) when the abuse allegedly occurred.

Information regarding the reporting of sexual violence, and the resources that are available to victims of sexual violence, is available at http://www.millersville.edu/socialeq/title-ix-sexual-misconduct/index.php.

Counseling Resources
Students sometimes face mental health or drug/alcohol challenges in their academic careers that interfere with their academic performance and goals. Millersville University is a caring community and resources are available to assist students who are dealing with problems. The Counseling Center (717-871-7821) is an important resource for both mental health and substance abuse issues. Additional resources include: Health Services (871-5250), Center for Health Education & Promotion (871-4141), Campus Ministries, and Learning Services (717-871-5554).

Course Web Site: http://cs.millersville.edu/~sschwartz/courses/csci-406-dm/