



Computer Science 493/579: DATA MINING AND KNOWLEDGE DISCOVERY
Fall 2013-14

Instructor: Dr. Sumeet Dua, Professor of Computer Science
E-mail: sdua@latech.edu
Office: Nethken -121
Instructor's Website: <http://www.latech.edu/~sdua>
Office Hours: M-F 8:00-10:00 a.m.

Text (Recommended): Data Mining: Concepts and Techniques
by Jiawei Han, Micheline Kamber and Jian Pei
Publisher: Morgan Kaufmann; 3rd edition

Reference texts:

Data Mining and Machine Learning for Cybersecurity
by Sumeet Dua and Xian Du
CRC Press; 1st edition, Hardcover.

Data Mining for Bioinformatics
by Sumeet Dua and Pradeep Chowriappa
CRC Press, 1st edition, Hardcover

Data Mining in Biomedical Imaging, Signaling and Systems
by Sumeet Dua and Rajendra Acharya U.
Auerbach Publishers, 1st edition, Hardcover.

In addition to the above, we will be using research articles, technical reports and published references for discussions and study.

Credit Hours: 3

Grade Distribution:

Final grades will be calculated as per the following distribution.

Activity	Percentage of Final Grade
Quizzes/Assignments	15%
Project	20% (proposal) + 25% (Final)= 45%
Final Exam	25%
Class Participation	15%
TOTAL (for CSC-493 enrollees)	100%
<i>Research Review (for CSC-579 enrollees)</i>	20%
TOTAL (for CSC-579 enrollees)	120%

Course Goals: To cover topics in data mining, including but not limited to, motivation for Data Mining (DM) motivation, knowledge discovery paradigm, preprocessing and normalization, dimensionality reduction, DM primitives, mining frequent itemsets, association rules, classification and evaluation measures. Differentiating, characterizing and evaluating supervised and unsupervised learning methods. Applications of data mining and knowledge discovery in databases in cybersecurity, biomedical, clinical and business applications.

Course Outcomes:

Upon successful completion of the course, the students will be able to:

- Understand the architecture of a typical data mining system.
- Understand various descriptive data summarization foundational techniques to measure tendency, dispersion and graphical inspection of high-dimensional data.
- Understand and implement data smoothing, Gaussian noise reduction and outlier identification and correction in streaming data.
- Understand and implement various normalization procedures including z-score and ratio-scaled normalizations.
- Understand Fourier transformations, discrete wavelet transformation and Parseval's theorem.
- Understand dimensionality reduction methods using Fourier transformation.
- Automatically generate concept hierarchies and discretize data using entropy based discretization and intuitive partitioning.
- Understand and implement algorithms for frequent itemset discovery [PO 3, 10, 12].
- Differentiate between levels, dimensions, correlation or associative and dense or sparse association rules.
- Define an association rule problem from variety of business and scientific problems.
- Understand differences between antimonotonic, monotonic, convertible, and inconvertible constraint-based rule mining.
- Understand and resolve issues with Classification and Prediction
- Understand decision-tree based classification
- Understand and employ Bayesian classification and rule-based classification
- Understand the intricacies of Associative Classification
- Understand and employ classification accuracy and error measures

GENERAL COURSE POLICIES

Moodle Policy: We will use Moodle for course documents and submissions. Ensure that you have access to Moodle and watch out for any announcements on there.

Late Assignment and Project Policy: Assignments, project and reports received after 24 hours of the expiry of due date and time, will be graded with a penalty of 50%. Assignments received after 24 hours of due date/time will be automatically assigned a grade of zero.

E-mail Policy: All course communications will be sent to your @latech.edu e-mail account. As a student of this course, you are responsible for checking your latech account at least once in 24 hours. Failure to read your e-mail regularly is not an excuse for delayed action.

No Plagiarism of programs, projects, reports, or exams will be excused and will lead to a failing grade in the course. Collaboration on assignments is limited to discussion of ideas and views. Collaborate intelligently.

Attendance: Attendance will be taken in class and late arrival in the class (after door closure) will be considered absenteeism. Do not miss classes without prior notification to the instructor. Class participation is important and absenteeism without excuses will lead to grade penalties.

Missed Activity Policy: No scheduled exam or course-activity must be missed without a university-approved excuse, pre-approved by the faculty. A grade of zero will be assigned for

missed activity. If the university-approved excuse for the course activity is accepted, the instructor will decide on the procedure for making up for that activity.

Academic Honor Code: Louisiana Tech University Honor Code is available at: <http://www.latech.edu/documents/honor-code.pdf>. In accordance with the Academic Honor Code, students pledge the following: Being a student of higher standards, I pledge to embody the principles of academic integrity.

Accommodations for students with disabilities: Students needing testing or classroom accommodations based on a disability are encouraged to discuss those needs with me as soon as possible. The students can also approach the Office of Disability Services (<http://www.latech.edu/ods>) for additional information.

Emergency Notification System (ENS): All Louisiana Tech students are strongly encouraged to enroll and update their contact information in the Emergency Notification System. It takes just a few seconds to ensure you're able to receive important text and voice alerts in the event of a campus emergency. For more information on the Emergency Notification System, please visit <http://www.latech.edu/administration/ens.shtml>.

Welcome to the course and have an enjoyable and informative quarter!