



**Computer Science 580: Advanced DATA MINING, FUSION AND APPLICATIONS**  
***Spring-2013***

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

**Text:**            **Data Mining and Machine Learning in Cybersecurity**  
By Sumeet Dua, Xian Du  
Publisher: CRC Press, 1<sup>st</sup> edition

**Data Mining: Concepts and Techniques** [selected topics]  
by Jiawei Han, Micheline Kamber  
Publisher: Morgan Kaufmann; 2<sup>nd</sup> edition

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.

<b>Activity</b>	<b>Percentage of Final Grade</b>
Project	15%(proposal)+25%(final) = 40%
Research Paper Review	20%
Final Exam	30%
Class Participation	10%
<b>TOTAL</b>	<b>100%</b>

Details on the project and assignments will be provided in the class.

**Course Objective:**

This course offering is to enable students of any discipline to learn about the principles of Data Mining and Data Fusion and its applications in Cyber Security and other data-rich areas. The course covers the fundamentals of High Dimensional data clustering • Types of cluster analysis and applications • Meta data detection • Correlation and Association Analysis • Resolution of semantic heterogeneity towards smooth distributed data integration • Implementation of Support Binary and multi-class classification and its application in intrusion detection • Application of spatio-temporal data structures for range queries for data mining applications • Intricacies of Image feature extraction for Content-Based Image Retrieval and Annotation of Images •

Advanced solutions in indexing and querying large time-series. *[New topics will be added/ substituted as needed]*

### **Course Outcomes:**

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

- Use data mining for financial data analysis, retail industry, telecommunications and scientific applications.
- Understand types of data for cluster analysis
- Perform clustering of high-dimensional data.
- Understand classical (K-means, hierarchical, Naïve Bayes) clustering algorithms and design new associative algorithms.
- Understand partitioning, hierarchical, density-based and grid-based clustering algorithms.
- Understand intricacies of metadata detection, correlation analysis, data conflict detection and the resolution of semantic heterogeneity towards smooth distributed data integration.
- Understand intricacies of image feature extraction for content-based image retrieval and annotation of images.
- Understand advanced solutions in indexing and querying large time-series.
- Understand the computational challenges in the analysis of available genomics and proteomics databases.
- Implement Support-vector machines algorithm for multi-class classification.
- Employ spatio-temporal data structures for range queries for data mining applications

### **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.

#### **Attendance Policy:**

Do not miss classes without prior notification to the instructor. Class participation is important in this course and absenteeism without excuses will lead to grade penalties over and above the above limits.

#### **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.

**Missed Activity Policy:**

No scheduled exam or course-activity can be missed without a university-approved excuse accepted by the instructor at least 24 hours before the scheduled activity. A grade of zero will be assigned for unexcused 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.

**General Course Policy:**

1. **NO PLAGIARISM** of programs, projects, reports, or exams will be excused and will lead to a failing grade in the course.
2. Collaboration on assignments is limited to discussion of ideas and views.
3. Late arrival in the class (after door closure) will be considered absenteeism

**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.

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