

Advanced Databases and Data Mining

Department of Computer Science
College of Science and Engineering
Texas A&M University – Commerce

Instructor: Truong-Huy D. Nguyen, Ph.D.

Assistant Professor

Email: truong-huy.nguyen@tamuc.edu

Classroom: EDS127

Meeting Times: Thursday, 4:30pm to 7:10pm

Office Hours: Thursday, 1:00pm to 4:00pm

COURSE OBJECTIVES

- Gain a working knowledge of data mining techniques
- Learn to design and implement algorithms to apply techniques in a practical fashion
- Understand which algorithms to apply to what kind of databases to obtain desired useful knowledge about the data

COURSE OUTLINE

General theory, concept, and techniques related to intelligent database design are discussed in this course. Many programming techniques to allow students to integrate intelligent database systems with web-based applications are covered in detail. Additional topics to be covered include expert systems, neural networks, hypermedia, and text retrieval. These additional topics are used as supplementary entities to be integrated into database reasoning components. A moderate-size semester project needs to be completed to demonstrate the ability to conduct data mining tasks on a realistic-sized database.

SOME TOPICS TO BE SELECTED FOR DISCUSSION

- Statistical Data Mining
- Classification
- Association Rules
- Clustering
- Web Mining
- Fuzzy Logic
- Rough Set Theory
- Neural Networks

TENTATIVE SCHEDULE

(1 Week) Chapter 1: Introduction
(1 Week) Chapter 2: Association Rules
(1 Week) Chapter 3: Classification Rules
(2 Weeks) Chapter 7: Clustering
(1 Week) Chapter 6: Neural Networks in Data Mining
(1 Week) Mid-term Test
(1 Week) Chapter 4: Statistical Data Mining
(2 Weeks) Chapter 5: Rough Sets and Data Mining
(1 Week) Chapter 8: Fuzzy Information Retrieval
(2 Weeks) Project and Demonstration

TENTATIVE RUBRIC

Homework (30%) – done in groups of up to 5
Mid-Term Test (30%)
Project Proposal (10%) – Project may be done in groups of up to 5 students
Project Performance (15%) – How much the team complete as compared to what initially proposed and approved
Project Presentation (15%)

TEXTBOOK

Practical Applications of Data Mining: Methods and Practices, Sang C. Suh, Jones and Bartlett Publishers, ISBN-13: 9780763785871, 2011.

ADDITIONAL USEFUL READING

[1]Bhavani Thuraisingham, Data Mining: Technologies, Techniques, Tools, and Trends, CRC Press (1999).
[2]Gordon S. Linoff and Michael J. A. Berry, Mining the Web, Wiley (2001).
[3]Tapio Eloma, Heikki Mannila, and Hannu Toivonen (Editors), Principles of Data Mining and Knowledge Discovery, Springer (2002).
[4]Stephan Kudyba and Richard Hoptroff, Data Mining and Business Intelligence: A Guide to Productivity, IDEA Group Publishing (2001).
[5]J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2000.

For research assistance, please contact:

John Atabaev, Reference Librarian: Sciences & Collection Development
John.Atabaev@tamuc.edu
903-886-5726

COURSE REQUIREMENTS

CSCI526 or equivalent is a prerequisite for this course. It is expected that each student be familiar with relational DBMS including relational data models, ER models, normalization, and functional dependencies. Though each of these will be reviewed in class during the first few weeks of the course, it is very imperative that a basic understanding of them exists. A significant amount of research is expected to be performed by each student to develop appropriate ideas and concepts relevant to intelligent database design. The main focus of this course will be on developing a working experience in data mining on a realistically sized database. In the second half of the semester, students are tasked to interact with an intelligent database system and apply concepts and ideas derived in class to propose and complete a course project. Each development project will be different in that each system should display a level of unique intelligent features identified throughout the course. Additional requirements of the course are a number of presentations to be made by each team, a term paper, and a project report. Due dates for all assigned materials will be announced in class in advance. It is the student's responsibility to have all assignments ready on time. Any student who has to be absent on an assignment due date must arrange to have the assignment submitted early. Late assignment may not be accepted.

➤ **ACADEMIC ETHICS:**

"All students enrolled at the University shall follow the tenets of common decency and acceptable behavior conducive to a positive learning environment." (See Student's Guide Handbook, Policies and Procedures, Conduct). Most importantly, I have zero tolerance for any cheating behavior. Students are immediately reported to the university should there be any cheating discovered. Cheating includes, but not limited to, plagiarism and peer solution copying.

➤ **ATTENDANCE POLICY:**

Students are expected to be present at all class lectures. If a student is absent from class on the due date of any assignment, they are expected to make alternative arrangements to assure that the assignment is turned in ON TIME. Any student wishing to withdraw from the course must do so officially as outlined in the class schedule. THE INSTRUCTOR CANNOT DROP OR WITHDRAW ANY STUDENT.

➤ **COURSE REQUIREMENT DEADLINES:**

Credit will be given for ONLY those exam(s), program(s), and/or project(s) turned in no later than the deadline(s) as announced by the instructor of this class unless prior arrangement has been made with the instructor.

➤ **SMOKE, VAPOR & TOBACCO FREE ENVIRONMENT:**

University Procedure 34.05.99.R1 now prohibits the use of vapor/electronic cigarettes, smokeless tobacco, snuff and chewing tobacco inside and adjacent to any building owned, leased, or operated by A&M – Commerce.

STUDENTS WITH DISABILITIES ACT COMPLIANCE

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you have a disability requiring an accommodation, please contact:

Office of Student Disability Resources and Services

Texas A&M University-Commerce

Gee Library

Room 132

Phone (903) 886-5150 or (903) 886-5835

Fax (903) 468-8148

StudentDisabilityServices@tamuc.edu