



Advanced Databases and Data Mining

Spring 2016

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

Instructor: Oner Celepcikay, Ph.D.

Email: Oner.Celepcikay@tamuc.edu (preferred communication method)

Phone: 713-966-9354

Classroom: Online

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 developed and demonstrated for the practice of the design of an intelligent database. This course project is a purely web-based and practical database system that is capable of handling high-level intelligent queries.

Textbook:

Introduction to Data Mining, by Pang-Ning Tan, Michael Steinbach and Vipin Kumar, 2006, 1st Edition, Addison-Wesley. Paper Book-ISBN 13: 978-0321321367

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 Hoptruff, 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.



Tentative Course Outline

#	Date	Topic
1	TBA	Course Introduction , Syllabus Overview What is Data Mining? DM Chapter 1
2	TBA	Data , DM Chapter: 2
3	TBA	Exploring Data , DM Chapter 3
4	TBA	Classification & Decision Trees , DM Chapter 4 - Measures of Node Impurity - Classification Issues
5	TBA	- Exam I - <i>R project (Classification)</i>
6	TBA	- Classification Model Evaluation - Comparing Classification Models
7	TBA	Cluster Analysis , DM Chapter 8 - Clustering Overview - K-means Clustering
		- Spring Break
8	TBA	- Hierarchical Clustering - DBSCAN Clustering - Cluster Evaluation
9	TBA	Cluster Analysis II - DM Chapter 9
10	TBA	- Exam 2 - <i>R project: Clustering</i>
11	TBA	- Exam 2 Review - Top 10 Data Mining Algorithms* - Google's PageRank*
12	TBA	- Regression Analysis (Subject to Change) - Association Analysis DM Ch 6 (Subject to Change)
13	TBA	- R vs. Python Projects
14	TBA	Final Exam

*These items are outside materials and subject to change depending on how much of the material we cover.



TENTATIVE RUBRIC:

Course grades will be determined as follows:

Activity	Weight
Exam-1 [DM Chapters 1, 2, 3, 4]	20%
Exam-2 [DM Chapter 4, 5, 8, 9]	20%
Final [All Covered Materials]	20%
Quizzes	5%
Homework Assignments	30%
Class Participation	5%

Your final course grade will be determined by following standard college formula:

90-100→"A", 80-89→"B", 70-79→"C", 60-69→"D", 0-59→"F"

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 three weeks of the course, it is very imperative that a basic understanding of them exists. During the first half of the semester, theoretical aspects of relational and intelligent databases will be covered. Most of this material will be derived from instructor's note, published papers, reference books, and class discussions. Therefore, significant amount of research is expected to be performed by each student to develop appropriate ideas and concepts relevant to intelligent database design. The second half of the semester will be used to develop an intelligent database system by applying concepts and ideas derived during the first half of the semester. Each development project will be different in that each system should display a level of unique intelligent features identified during the first half of the course. Additional requirements of the course are a number of presentations to be made by each team, a test, 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).



➤ **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