COUESE OUTLINE

DATA MINING
Spring 2025
COSC 4352.001

Tuesday-Thursday: 2:00-3:20 PM in COB 211

Instructor: Arun Kulkarni, Ph.D.,
Professor of Computer Science

Office: COB 315.07
Office Hours: T-R: 12:30-2:00 PM
Email: akulkarni@uttyler.edu

Textbook: Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar (2019). Introduction to Data Mining, 2nd Edition, Pearson, New York, NY

Additional Book: Jiawei Han and Micheline Kamber, Jian Pei (2011). *Data Mining – Concepts and Techniques*. Morgan Kaufmann, San Francisco, CA

Our ability to capture and store data has far outpaced our ability to process and utilize it. Business data contain valuable information. The course deals with knowledge discovery from databases (KDD). Data Mining represents a sub set of a KDD process. Data Mining involves model fitting, classification, estimation, prediction, and clustering. In this course, we will discuss various stages in a KDD system, algorithms for data mining, pre-processing techniques, and algorithms for extracting knowledge from data models. Tentative topics are described below.

Course Objectives

1. **Explain** the concepts, goals, and applications of data mining and its relationship to related fields.
2. **Analyze** data characteristics and quality using exploratory techniques to understand underlying patterns and issues.
3. **Apply** data preprocessing methods such as cleaning, transformation, and reduction to prepare data for mining tasks.
4. **Implement** frequent pattern and association rule mining techniques, including the Apriori algorithm, to discover meaningful relationships.
5. **Apply and compare** clustering techniques such as Hierarchical Clustering and K-Means to group data and assess cluster quality.
6. **Create and Apply** classification models using decision trees and neural networks for predictive analysis.

7. **Design and interpret** fuzzy inference systems and knowledge extraction processes to derive actionable insights from data. *(Create / Evaluate)*

TOPICS	HOURS
Introduction to data mining	3
Getting to know your data	3
Pre-processing techniques	6
Mining frequent patterns, association, a priori algorithm.	6
Clustering: Hierarchical Clustering, K-Means clustering	9
Classification Basic concepts: decision trees	6
Neural networks	3
Fuzzy Inference Systems	3
Knowledge extraction	3

Evaluation

Test 1	Thursday, 19-FEB-2026	70 %
Test 2	Thursday, 9-APR-2026	
Final examination	Tuesday, 28-APR-2026	
Assignments		20 %
Attendance & Class Participation		10 %

Academic Dishonesty: You are expected to do your own work. You may assist each other with general concepts, but direct assistance with a particular assignment or any attempts to gain an unfair academic advantage will not be tolerated. Cheating is considered a serious academic offense both by the department and the University. It may result in a failing grade from this course for all parties involved. If you have questions about the line between assistance and cheating, discuss it with your instructor. The instructor reserves the right to ask you to explain any assignment that you turn in to judge if the work is actually yours.

Disabilities: If you have a disability, including a learning disability, for which you request an accommodation, please contact the Disability Support Services office so that the appropriate arrangements may be made. In accordance with federal law, a student requesting accommodation must provide documentation of his/her disability to the Disability Support Services counselor. For more information, call or visit the Student Services Center located in the University Center, Room 282. The telephone number is 566-7079 (TDD 565-5579).