

The University of Texas at El Paso
Department of Computer Science
Course Title: Data Mining
Course No.: CS 4362 (CRN: 23700), CS 5362 (CRN: 23726)
Spring 2023 Syllabus

Course description: Data mining refers to data exploration to discover knowledge in the form of patterns and structures. The knowledge found goes beyond the general pattern search where queries are known. Data mining is an analytic process to discover unknowns about data. The topics covered in this course help gain insights about data and develop expertise in large-scale data mining. Classroom activities include exercises along with regular lectures and discussions.

Class meetings: Mondays and Wednesdays 12:00pm-1:20pm

Location: Liberal Arts Building, Room 108

Instructor: Dr. Mahmud Shahriar Hossain

Email: mhossain [at] utep [dot] edu

Web: <https://www.cs.utep.edu/mhossain/>

Office Hours: Mondays and Wednesdays 1:45 PM – 2:45 PM.

Office location: Chemistry and Computer Sci Bldg 3.0504

Prerequisite for the course: C or better in Data structures for CS 4362.

Reference books:

- Introduction to Data Science: Computing For All:
<https://computing4all.com/courses/introductory-data-science/>
- Data Mining: Concepts and Techniques, 3rd Edition, Jiawei Han, Micheline Kamber, Jian Pei
- Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Ian Witten, Eibe Frank, Mark Hall
- Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Vipin Kumar
- The Elements of Statistical Learning, 2nd Edition, Trevor Hastie, Robert Tibshirani, Jerome Friedman, <https://web.stanford.edu/~hastie/Papers/ESLII.pdf>
- Mining of Massive Datasets, Anand Rajaraman, Jure Leskovec, Jeffrey D. Ullman, <http://infolab.stanford.edu/~ullman/mmds/book.pdf>
- Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, <http://nlp.stanford.edu/IR-book/>

OUTCOMES:

On successful completion of this course, students will gain the following expertise.

1. Students will be able to connect data science problems with algorithms or a family of algorithms.
2. Students will gain expertise in open-form data exploration in absence of a given problem.
3. Students will become familiar with the data mining application development process.

GRADUATE LEVEL VS. UNDERGRADUATE LEVEL EXPECTATIONS:

Graduate students will have additional tasks in some assignments/homework.

EVALUATION:

Grading components:

Midterm	30%
Final	20%
Quiz/ in-class exercise	10%
Homework	35%
Attendance	5%

The grading scale is A: 90-100, B: 80-89, C: 70-79, D: 60-69, F: below 60.

Grade appeals: All exam/homework/quiz grades must be appealed within 7 days of the grade being posted.

Exams: Two to three midterms and a final exam will be given. Make-up exams will not be given. Any student who misses an exam due to an unexcused absence will receive a grade of zero for that exam with no opportunity for make-up or substitution. University excused absences will be excused; exam-related arrangements should be made in advance in those cases. Military students are asked to inform the instructor about any anticipated absence as soon as possible.

Homework: Regular homework will be assigned which will require significant effort outside of class. The assignments are designed to challenge you by requiring that you apply learned concepts to new situations. You should start your homework immediately after you receive it.

Quizzes and exercises: There will be regular quizzes and exercises in the class. The quizzes and the in-class activities are not scheduled instead may appear suddenly any day.

Attendance: The instructor's policy is to penalize those students who are absent. Students are expected to participate in classes actively and show courtesy by not arriving late or leaving early. Although attendance has a specific weight, the instructor reserves the right to penalize the final grade for low attendance because the core to success is consistent and diligent effort. Attending classes is the minimum training for consistency.

CLASS REQUIREMENT:

Students must keep a laptop during the classroom sessions for in-class exercises and hands-on activities. We will be using Python for some of the exercises. It is recommended that students have Python installed in their computers.

CLASS POLICIES:

Standards of Conduct: In the classroom and in all academic activities, students are expected to uphold the highest standards of academic integrity. Any form of scholastic dishonesty is an affront to the pursuit of knowledge and jeopardizes the quality of the degree awarded to all graduates of UTEP.

Any student who commits an act of scholastic dishonesty is subject to discipline. Scholastic dishonesty includes, but is not limited to, cheating, plagiarism, uploading course materials or assignment questions to public forums or websites like chegg, the submission for credit of any work or any materials that are attributable in whole or in part to another person, taking an examination for another person, an act designed to give unfair advantage to a student or the attempt to commit such acts. Proven violations of the detailed regulations, as printed in the Handbook of Operating Procedures may result in sanctions ranging from disciplinary probation to a failing grade in the course, to suspension or dismissal, among others. The Handbook of Operating Procedures: Student Conduct and Discipline can be accessed at the following link: <https://www.utep.edu/hoop/section-2/student-conduct-and-discipline.html> .

DISABILITIES:

The University is committed to providing reasonable accommodations and auxiliary services to students, staff, faculty, job applicants, applicants for admissions, and other beneficiaries of University programs, services and activities with documented disabilities in order to provide them with equal opportunities to participate in programs, services, and activities in compliance with sections 503 and 504 of the Rehabilitation Act of 1973, as amended, and the Americans with Disabilities Act (ADA) of 1990 and the Americans with Disabilities Act Amendments Act (ADAAA) of 2008. Reasonable accommodations will be made unless it is determined that doing so would cause undue hardship on the University. Students requesting an accommodation based on a disability must register with the [UTEP Center for Accommodations and Support Services \(CASS\)](#). Contact the Center for Accommodations and Support Services at 915-747-5148, or email them at cass@utep.edu, or apply for accommodations online via the [CASS portal](#).

TOPICS:

In this course, students will gain hands-on experience on the data mining and knowledge discovery process. The topics covered in this course are as follows.

1. Data exploration
2. Similarity and distance measures
3. Cluster analysis
4. Regression-based prediction and autoregression
5. Classification
6. Pattern mining
7. PageRank algorithm
8. Outlier/anomaly detection
9. Mathematical optimization

Some additional items the instructor plans to include, but are subject to availability of time, are as follows.

10. Dimensionality reduction
11. Process-based parallelism (multiprocessing)
12. Time series analysis

The instructor will include topics that he finds relevant as the semester progresses.

The instructor reserves the right to make necessary changes to this syllabus and to the delivery of the course.