

The University of Texas at El Paso
Department of Computer Science
Course Title: Data Mining
Course No.: CS 4362 (CRN 17558)
Fall 2020 Syllabus

Course description: Data mining refers to the exploration of data to discover knowledge in the form of patterns and structures. The knowledge discovered 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 are useful to gain insights about data and develop expertise in large-scale data mining.

Along with regular lectures and discussions, there will be hands-on activities, especially on algorithm design and data analysis. Additionally, the course will cover recent state-of-the-art algorithms used to discover useful information from a massive amount of data. This course focuses on connecting theory and real-world data science problems.

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

Class meetings:

CS 4362: MW 10:30am-11:50am

Location: Blackboard Collaborate Ultra (available in the Blackboard entry for this course.)

Instructor: Mahmud Shahriar Hossain

Email: mhossain@utep.edu

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

Office hours: Mondays Wednesdays 2:00 PM – 3:00 PM and by appointment.

Office location: MS Teams ([Link to the MS Teams office room](#))

Teaching assistant: Ahnaf Farhan

Email: afarhan@miners.utep.edu

Office hours: Mondays and Wednesdays 12:00pm - 1:30pm, Thursdays 12:00pm - 1pm

Office location: Blackboard Collaborate Ultra (available in the Blackboard entry for this course.)

Reference books:

- 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.
4. Students will be able to design objective functions with free variables.
5. Students will become familiar with the use of multiple processors to make computation faster.

EVALUATION:

Grading components for undergraduate students:

Midterm	25%
Final	25%
Homework, quiz, in-class exercise	45%
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 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.

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 are not scheduled instead may appear suddenly any day. There may be individual exercises in the class, as well.

Attendance: The instructor's policy is to penalize those students who are absent. Students are expected to actively participate in classes and show the 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 based on the fact that active participation in the classroom environment is the core of this course.**

CLASS POLICIES:

It is recommended that you have a laptop during the synchronous class.

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, 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:

If you have a disability and need classroom accommodations, please contact The Center for Accommodations and Support Services (CASS) at 747-5148, or by email to cass@utep.edu. For additional information, please visit the CASS website at www.sa.utep.edu/cass.

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. Classification
3. Feature analysis
4. Association analysis
5. Link analysis
6. Cluster analysis
7. Graph mining
8. Anomaly detection
9. Visual analytics
10. Parallel computing

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

OTHER REQUIREMENTS: Students are encouraged to keep a laptop during the virtual classroom session 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 machines.

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