CSE 557 – Course Syllabus

Course No. & Title: CSE 557 – Data Mining

Term & Meeting Info: Fall Term, 2022-23 Tuesday 13:00-16:00

Instructor Info: Dr. Burcu YILMAZ, E-mail: byilmaz@gtu.edu.tr, Office: 235

Phone: (262) 605 22 24

Instructor Info: The course materials and announcements will be posted on Teams.

Catalog Description: This course provides an introduction to the fundamental principles and techniques of data mining. Theory is put into practice by investigating a variety of applications. Lectures help on how to solve real-life problems using machine learning methods.

Text Info: ş

- Data Mining Concepts and Techniques, Jiawei Han and Micheline Kamber, Third Edition, 2011
- Introduction to Data Mining, Tan, Steinbach, Kumar, First Edition, 2014, Pearson
- Introduction to Machine Learning, Ethem Alpaydin, Second Edition, October 2010, The MIT Press, (ISBN-10: 0-262-01211-1)

Slides: https://hanj.cs.illinois.edu/bk3/bk3_slidesindex.htm

Test & Grading Info:

Assignments (3 assignments)	15%
Final Project	25%
Midterm	25%
Final Exam	35 %

These weight are decided for the situation, where there will be online exam. Weights of the assignments and exams may change during the semester, if the exams are decided to be online. You will be informed about the changes.

Tentative Course Schedule:

Topics

- Introduction
- Exploring Data & Data Preprocessing
- Mining Frequent Patterns, Associations and Correlations
- Classification
- Classification (Advanced Techniques)
- Cluster Analysis
- Dimensionality Reduction
- Anamoly Detection
- Graph Mining Related Topics

- Text Mining
- Datawarehouses

Lectures: Lectures will be both face to face and online at the same time. There will be no recordings of lectures. Do not ask for it.

Attendance: You must attend at least %50 of the lectures.

Assignments: Assignments may have demo sections. If you do not attend the demo section of an assignment, you will be graded 0 for that assignment.

Final Project: In your project, you are obligated to select a topic related to data mining.

Rules:

- You should get an approval (face to face) from the lecturer about the topic selection to continue on the procedure. You may not get an approval just by e-mails.
- If you choose a topic that is not mentioned in the class, you may get higher grades than the other projects. You can not use knn or k-means in your projects. You can not use a technique that is given as an assignment. If you violate the rules, you will be graded 0.
- Other than implementing your project, you have to read 3 academic well known international papers about the project and prepare at least 1 page long summary of each paper. Do not send the papers by email.
- You have to attend demo section, prepare a presentation and an academic paper related to your project.
- In your presentation you should give the problem defintion, details related to your project (the data mining model and the evaluation results). **Do not present your thesis or projects from other courses!!! Addition to presentation,** there will be demo session. If you do not attend to demo session you will get "0" from your final Project.

Assignment Dates:

If you violate the due dates, you will get -2.5 points for every delayed day. Your homework will not be graded, if you delay your homework more than 20 days.

Additional Course Features:

Cheating: Cheating in this course is defined as the use of unauthorized material during the exam. Unauthorized material is defined as any material other than the exam sheets that will be provided throughout the semester and a calculator. Students identified as cheating will receive a grade of zero on the exam and/or a failing grade in the course. For projects, make sure everything is in your own words. If you have taken any discussion and/or excerpt from some outside sources, please make sure you give reference to it right after the discussion; otherwise you will get an automatic -100 from the project depending on the amount of unreferenced text or copy paste code from internet.

Schedule: Some variation from the above topical coverage and/or scheduling may occur. Also, some additional topics may be included. Notification of these will be given by the instructor during the lecture period. The student is responsible for all material covered in lecture.