COURSE TITLE/SECTION: Data Mining (COSC 6335)   *December 6, 2023*

TIME: TUTH 2:30-4p in S120

FACULTY: Christoph F. Eick   **OFFICE HOURS**: TU 4:10-5p TH 8:50-10a (MS Teams)

E-mail: ceick@uh.edu   **Phone: 33345 (use e-mail!!)**   **FAX: 33335**

Course Website: [COSC 6335: a Data Mining Course (Fall 2023) (uh.edu)](#)

Consult the course website for more information about COSC 6335! Use the course webpage (and not the MS Team page associated with the course) to download course information, such as lecture slides and read the news section regularly.

I.   Course *Data Mining (COSC 6335)*

A.   Catalog Description

Goals and objectives of data mining, data quality, data preprocessing, OLAP and data warehousing, exploratory data analysis, classification and prediction, similarity assessment, cluster and outlier analysis, association analysis, post processing techniques, data mining methodologies, data mining case studies.

B.   Purpose

Data mining centers on finding novel, interesting, and potentially useful patterns in data. It aims at transforming a large amount of data into a *well of knowledge*. Data mining has become a very important field in industry as well as academia. The course covers most of the important data mining techniques and provides background knowledge on how to conduct a data mining project. Topics covered in the course include exploratory data analysis, classification and prediction, clustering and similarity assessment, association analysis, outlier and anomaly detection, and interpreting and interpreting data analysis/data mining results. In the first 9 weeks a very basic introduction to data mining will be given. After defining what knowledge discovery and data mining is, data mining tasks such classification, clustering, and association analysis will be discussed in detail. Also basic visualization techniques and statistical methods will be introduced. Moreover, hands on data mining experience will be provided in three problem sets. Finally, you will learn on how to use and do programming in the popular statistics, visualization, and data mining environment *R*.

II.   Course Objectives

Upon completion of this course, students
1.   will know what the goals and objectives of data mining are and how to conduct a data mining project
2.   will have sound knowledge of popular classification techniques, such as decision trees, support vector machines and neural networks.
3.   will know the most important association analysis techniques
4.   will have detailed knowledge of popular clustering algorithms such as K-means, density-based, graph-based, hierarchical clustering and cluster evaluation.
5.   will obtain some basic knowledge about popular outlier detection techniques
6.   will conduct small and medium-sized projects in which data mining is applied to real world data sets. They will obtain valuable experience in learning how to interpret and evaluate data mining results, how to select parameters of data mining tools, and how to make sense out of data.
7.   will get some practical experience in evaluating data mining results of other students in the course as well as data mining publications. Kritik (https://www.kritik.io/) will be used for some evaluation tasks of the course.
8.   will obtain practical experience in designing and implementing data mining algorithms
9.   will learn on how to use popular data mining programming environment **R** and/or Python Data Science Libraries

III.   Course Organization

I.   Introduction to Data Mining
II.   Data Science Basics and Exploratory Data Analysis
III.   Brief Introduction to Peer Reviewing and using Kritik for it
IV.   Lab: Using R and Python for Data Science and Data Mining
V.   Data Storytelling
VI.   Introduction to Clustering and Similarity Assessment
VII.   Density Estimation
VIII.   Outlier and Anomaly Detection
IX.   Introduction to Classification: Basic Concepts and Decision Trees, Support Vector Machines and Neural Networks.
X.   Introduction to Deep Learning Centering on Autoencoders
XI.   Reviewing Data Mining Papers
XII.   Spatial Data Mining (optional topic)
XIII.   Association Analysis, centering on APRIORI
XIV.   Data Preprocessing
XV.   Advanced Clustering

IV.   Course Structure

23 lectures
2 exams (Oct. 19 and Dec. 7, 2023)
3 problem sets (some problem set tasks are group tasks!)
1 student presentation as part as a group homework credit task[1]
2 40-minute review sessions

V.   Problem Sets

Problem Sets contain paper and pencil tasks which review your understanding of basic data mining concepts and algorithms, tasks which use data mining tools, and small and medium sized data analysis/data mining projects, and tasks in which you evaluate data mining results of other students and data mining publications. Some tasks will be group tasks. There will be three Problem Sets in Fall 2023:

Problem Set1: Exploratory Data Analysis and Development of a Data Storytelling Tool
Problem Set2: Clustering, Outlier Detection and Density Estimation
Problem Set3: Data Mining Paper Reviewing

---

[1] Groups of students solve homework-style problems and present their solution during the lecture!

VI.   Textbooks

Highly Recommended Text:
> P.-N. Tang, M. Steinback, and V. Kumar *Introduction to Data Mining*,
> Addison Wesley, Second Edition.

Recommended Text:
> Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*
> Morgan Kaufman Publishers, Third Edition

VII.   Evaluation and Grading
Remark: subject to change; check course website again by Sept. 9, 2023 for finalized information.

Problem Sets: 47% (Task Weights: T1:4, T2:4, T3:4, T4:5, T5:1.5)
Group Homework Credit: 3%
Midterm Exam: 21%
Final Exam: 27%
Attendance: 2%

Course exam will be open book/notes paper exams, but the use of cell phones and computers is not allowed during the exams.

Students will be responsible for material covered in the lectures and assigned in the readings.

Translation number to letter grades:
A:100-90 A-:90-85 B+:85-82 B:82-77 B-:77-74 C+:74-70
C: 70-66 C-:66-62 D+:62-58 D:58-54 D-:54-50 F: 50-0

Students may discuss course material and homeworks, but must take special care to discern the difference between **collaborating** in order to increase understanding of course materials and collaborating on the homework / course project itself. We encourage students to help each other understand course material to clarify the meaning of homework problems or to discuss problem solving strategies, but it is **not** permissible for one student to help or be helped by another student in working through homework problems and in the course project. If, in discussing course materials and problems, students believe that their like-mindedness from such discussions could be construed as collaboration on their assignments, students must cite each other, briefly explaining the extent of their collaboration. Any assistance that is not given proper citation may be considered a violation of the Honor Code, and might result in obtaining a grade of F in the course, and in further prosecution.

> **Policy on grades of I (Incomplete):** A grade of 'I' will only be given in extreme emergency situations and only if the student completed more than 2/3 of the course work.

VIII. Course Exams

Course exams will be paper exams which you take in a UH classroom, likely our assigned classroom. If you are experiencing any COVID-19 symptoms that are not clearly related to a pre-existing medical condition, do not come to the exam. Please see Student Protocols for what to do if you experience symptoms and Potential Exposure to Coronavirus for what to do if you have potentially been exposed to COVID-19. Consult the (Graduate Excused Absence Policy) for information regarding excused absences due to medical reasons. Exams will be open books/notes paper exams, but the use of cell phones and computers is not allowed.

Moreover, if you miss a course exam for reasons that are not covered by UH Excused Absence policies you will get a grade of 'F' for the missed exam.

IX. Mental Health and Wellness Resources

The University of Houston has a number of resources to support students' mental health and overall wellness, including CoogsCARE and the UH Go App. UH Counseling and Psychological Services (CAPS) offers 24/7 mental health support for all students, addressing various concerns like stress, college adjustment and sadness. CAPS provides individual and couples counseling, group therapy, workshops and connections to other support services on and off-campus. For assistance visit uh.edu/caps, call 713-743-5454, or visit a Let's Talk location in-person or virtually. Let's Talk are daily, informal confidential consultations with CAPS therapists where no appointment or paperwork is needed. The Student Health Center offers a Psychiatry Clinic for enrolled UH students. Call 713-743-5149 during clinic hours, Monday through Friday 8 a.m. - 4:30 p.m. to schedule an appointment. The A.D. Bruce Religion Center offers spiritual support and a variety of programs centered on well-being.

Need Support Now?

If you or someone you know is struggling or in crisis, help is available. Call CAPS crisis support 24/7 at 713-743-5454, or the National Suicide and Crisis Lifeline: call or text 988, or chat 988lifeline.org.

X. <u>Title IX/Sexual Misconduct</u>

Per the UHS Sexual Misconduct Policy, your instructor is a "responsible employee" for reporting purposes under Title IX regulations and state law and must report incidents of sexual misconduct (sexual harassment, non-consensual sexual contact, sexual assault, sexual exploitation, sexual intimidation, intimate partner violence, or stalking) about which they become aware to the Title IX office. Please know there are places on campus where you can make a report in confidence. You can find more information about resources on the Title IX website at https://uh.edu/equal-opportunity/title-ix-sexual-misconduct/resources/.

XI. Reasonable Academic Adjustments/Auxiliary Aids

The University of Houston complies with Section 504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990, pertaining to the provision of reasonable academic adjustments/auxiliary aids for disabled students. In accordance with Section 504 and ADA guidelines, UH strives to provide reasonable academic adjustments/auxiliary aids to students who request and require them. If you believe that you have a disability requiring an academic adjustments/auxiliary aid, please contact the Justin Dart Jr. Student Accessibility Center (formerly the Justin Dart, Jr. Center for Students with Disabilities).

XII. Excused Absence Policy

Regular class attendance, participation, and engagement in coursework are important contributors to student success. Absences may be excused as provided in the University of Houston Undergraduate Excused Absence Policy and Graduate Excused Absence Policy for reasons including: medical illness of student or close relative, death of a close family member, legal or government proceeding that a student is obligated to attend, recognized professional and educational activities where the student is presenting, and University-sponsored activity or athletic competition. Under these policies, students with excused absences will be provided with an opportunity to make up any quiz, exam or other work that contributes to the course grade or a satisfactory alternative. Please read the full policy for details regarding reasons for excused absences, the approval process, and extended absences. Additional policies address absences related to military service, religious holy days, pregnancy and related conditions, and disability.

XIII. Recording of Class

Students may not record all or parts of class, livestream all or part of class, or make/distribute screen captures, without advanced written consent of the instructor. If you have or think you may have a disability such that you need to record class-related activities, please contact the Justin Dart, Jr. Student Accessibility Center. If you have an accommodation to record class-related activities, those recordings may not be shared with any other student, whether in this course or not, or with any other person or on any other platform. Classes may be recorded by the instructor. Students may use instructor's recordings for their own studying and notetaking. Instructor's recordings are not authorized to be shared with *anyone* without the prior written approval of the instructor. Failure to comply with requirements regarding recordings will result in a disciplinary referral to the Dean of Students Office and may result in disciplinary action.

XIV. Syllabus Changes

Please note that the instructor may need to make modifications to the course syllabus and may do so at any time. Notice of such changes will be announced as quickly as possible through the course webpage!

XV. Resources for Online Learning

The University of Houston is committed to student success, and provides information to optimize the online learning experience through our Power-On website. Please visit this website for a comprehensive set of resources, tools, and tips including: obtaining access to the internet, AccessUH, and Blackboard; requesting a laptop through the Laptop Loaner Program; using your smartphone as a webcam; and x If you are experiencing any COVID-19 symptoms that are not clearly related to a pre-existing medical condition, do not come to class. Please see Student Protocols for what to do if you experience symptoms and Potential Exposure to Coronavirus for what to do if you have potentially been exposed to COVID-19. Consult the (select: Undergraduate Excused Absence Policy or Graduate Excused Absence Policy) for information regarding excused absences due to medical reasons.

XVI. UH Email

Please check and use your Cougarnet email for communications related to this course. You will also need your Cougarnet account to access course MS Teams page 3337-Class and for listening to course lecture. To access this email, login to your Microsoft 365 account with your Cougarnet credentials. Your Cougarnet Account also allows you to download Microsoft Office 365 at no cost. For questions or assistance contact UHOnline@uh.edu.

XVII. Academic Honesty Policy

High ethical standards are critical to the integrity of any institution, and bear directly on the ultimate value of conferred degrees. All UH community members are expected to contribute to an atmosphere of the highest possible ethical standards. Maintaining such an atmosphere requires that any instances of academic dishonesty be recognized and addressed. The UH Academic Honesty Policy is designed to handle those instances with fairness to all parties involved: the students, the instructors, and the University itself. All students and faculty of the University of Houston are responsible for being familiar with this policy.

XVIII. Security Escorts and Cougar Ride

UHPD continually works with the University community to make the campus a safe place to learn, work, and live. Our Security escort service is designed for the community members who have safety concerns and would like to have a Security Officer walk with them, for their safety, as they make their way across campus. Based on availability either a UHPD Security Officer or Police Officer will escort students, faculty, and staff to locations beginning and ending on campus. If you feel that you need a Security Officer to walk with you for your safety please call 713-743-3333. Arrangements may be made for special needs.

Parking and Transportation Services also offers a late-night, on-demand shuttle service called Cougar Ride that provides rides to and from all on-campus shuttle stops, as well as the MD Anderson Library, Cougar Village/Moody Towers and the UH Technology Bridge. Rides can be requested through the UH Go app. Days and hours of operation can be found at https://uh.edu/af-university-services/parking/cougar-ride/.

XIX. Use of AI related and Other Tools

The use of AI-related tools, such ChatGPT, is permitted in this course. However, student reports have to cite all AI-related tools and other tools and documents that were used to conduct the task at hand and for what they were used. Not citing used tools and documents in student reports, represents a serious academic honesty violation, and violators will be prosecuted.

XX Other Helpful Links

Coogs Care: https://uh.edu/dsa/coogscare/

Student Health Center: https://www.uh.edu/healthcenter/