UNIVERSITY of HOUSTON

COLLEGE OF NATURAL SCIENCES & MATHEMATICS HTTP://NSM.UH.EDU

COURSE TITLE/SECTION: Data Science I (COSC 3337) *November 22, 2021*

TIME: TT 1:30a-1p

FACULTY: Christoph F. Eick **OFFICE HOURS**: MO 4-4:45p WE 9-10a
E-mail: ceick@uh.edu **Phone: 33345 (use e-mail!!)** **FAX: 33335**

I. Course *Data Science I (COSC 3337)*

Credit Hours: 3.0

Lecture Contact Hours: 3 *Lab Contact Hours:* 0

Formerly COSC 4335.
Prerequisite: A grade of C- or better in COSC 2306 or COSC 2436, and MATH 3339 and declared COSC major, COSC minor, or Data Science minor.
Description
Data science concepts including exploratory data analysis, data visualization, statistical inference and modeling, machine learning, clustering, post-processing and interpreting results.

II. Course Objectives

Upon completion of this course, students

1. will know what the goals and objectives of data science are and how to conduct a data science project
2. will have a sound knowledge of basic statistics and basic machine learning concepts
3. will have some sound knowledge about exploratory data analysis and data visualization techniques used in exploratory data analysis
4. will have knowledge of popular classification techniques, such as decision trees, support vector machines, ensembles, and neural networks
5. will have some sound knowledge about how to construct distance functions
6. will have detailed knowledge of popular clustering algorithms, such as K-means, DBSCAN, and hierarchical clustering and cluster evaluation.
7. will get hands-on exposure in the course problem sets and group project how to apply data analysis techniques to real world data sets. They will obtain valuable experience in learning how to interpret data analysis results, how to select parameters of data analysis tools, and data story telling
8. will learn on how to use the data analysis and visualization environments such as **R** and its popular libraries and how to develop software on the top of R.
9. will have some basic knowledge concerning outlier detection and association analysis.

III. Course Content

1. Introduction to Data Analysis
2. Exploratory Data Analysis—how to Visualize and Compute Basic Statistics for Datasets and How to Interpret the Findings
3. Brief Introduction to R (optional topic)
4. Introduction to Supervised Learning: Basic Concepts and Decision Trees
5. More on Supervised Learning: Instance-based Learning, Support Vector Machines, Neural Networks, Regression
6. Similarity Assessment—how to Obtain Distance Functions
7. Introduction to Clustering
8. Anomaly and Outlier Detection
9. Association Analysis
10. Data Storytelling

There will be 3 Problem Sets each consisting of individual tasks, centering on: Problem Set1: Exploratory Data Analysis and Data Storytelling
Problem Set2: Hyperparamter Selection for Neural Networks
Problem Set3: Clustering

There will be an 7-week long group project (3-4 students per group, September 24 to November 12) in which groups: a. identify a dataset b. determine what question(s) should be answered using this dataset c. employ ML/Statistics/DM techniques to answer the identified question(s) d. summarize their findings in a report and a short presentation.

IV. Course Structure

22 lectures
1-2 labs
2 exams
3 problem sets
1 group project

2 presentations (group project and group homwork credit in which groups of students present solutions to home-work style problems (also see food note)

V. Textbooks

Recommended Text:

P.-N. Tang, M. Steinback, and V. Kumar *Introduction to Data Mining*, Addison Wesley, 2018.

Maybe:

Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from the Frontline. O' Reilly. 2014

VII. Evaluation and Grading

Problem Sets: 33% , Group Project: 16% and Group Homework Credit:3%
Exams: 48% (Midterm Exam: 20%, Final Exam: 28%)

Translation number to letter grades:
A:100-92 A-:92-88 B+:88-84 B:84-80 B-:80-76 C+:76-71
C: 71-66 C-:66-62 D+:62-58 D:58-54 D-:54-50 F: 50-0

Students may discuss course material and assignments 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.

UH Undergraduate Excused Absence Policy applies for missing exams; students who miss exams for reasons, not stated in this policy, will be a grade of 'F' in the exam.

VIII. Consultation

Instructor: Dr. Christoph F. Eick
office hours (MS Teams for the time being): TU 1-2p TH 9:30-10:30a
e-mail: ceick@uh.edu
class meets: TU/TH 11:30a-1p

IX. Bibliography

The following conferences and journals center on data science and related areas:

1. Data mining and KDD
    - Conference proceedings: ICDM, KDD, PKDD, PAKDD, etc.

1 *Online Credit*: Different Homework-style problems will be assigned to groups of about 4 students (each group gets a different task) and each group presents their solutions in a presentation online (about 10 minutes) and shares them by providing a Word/pptx file. This assumes that the course will be partially taught online; if less than 20% of the course will be taught online, online credit tasks will be dropped from the course content.

- Journal: Data Mining and Knowledge Discovery

2. Database field (SIGMOD member CD ROM):
    - Conference proceedings: VLDB, ICDE, ACM-SIGMOD, CIKM
    - Journals: ACM-TODS, J. ACM, IEEE-TKDE, JIIS, etc.
3. AI and Machine Learning:
    - Conference proceedings: ICML, AAAI, IJCAI, etc.
    - Journals: Machine Learning, Artificial Intelligence, etc.
4. Statistics:
    - Conference proceedings: Joint Stat. Meeting, etc.
    - Journals: Annals of statistics, etc.
5. Visualization:
    - Conference proceedings: CHI, etc.
    - Journals: IEEE Trans. visualization and computer graphics, etc.

Face Covering Policy

To reduce the spread of COVID-19, the University strongly encourages everyone (vaccinated or not) to wear face coverings indoors on campus including classrooms for both faculty and students.

Presence in Class

Your presence in class each session means that you:

- Are NOT exhibiting any Coronavirus Symptoms that makes you think that you may have COVID-19
- Have NOT tested positive or been diagnosed for COVID-19
- Have NOT knowingly been exposed to someone with COVID-19 or suspected/presumed COVID-19

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.

COVID-19 Information

Students are encouraged to visit the University's COVID-19 website for important information including on-campus testing, vaccines, diagnosis and symptom protocols, campus cleaning and safety practices, report forms, and

positive cases on campus. Please check the website throughout the semester for updates.

Vaccinations

Data suggests that vaccination remains the best intervention for reliable protection against COVID-19. Students are asked to familiarize themselves with pertinent vaccine information, consult with their health care provider. The University strongly encourages all students, faculty and staff to be vaccinated.

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).

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.

Recording of Class

Students may not record all or part 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.

Syllabus Changes

Due to the changing nature of the COVID-19 pandemic, 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 on the course webpage in the News Section.

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 downloading Microsoft Office 365 at no cost. For questions or assistance contact UHOnline@uh.edu.

UH Email

Please check and use your Cougarnet email for communications related to this course. To access this email, login to your Microsoft 365 account with your Cougarnet credentials.

Webcams

Access to a webcam is required for students participating remotely in this course. Webcams must be turned on (*state when webcams are required to be on and the academic basis for requiring them to be on*). (*Example: Webcams must be turned on during exams to ensure the academic integrity of exam administration*.)

Honor Code Statement

Students may be asked to sign an honor code statement as part of their submission of any graded work including but not limited to projects, quizzes, and exams: " *I understand and agree to abide by the provisions in the (select: University of Houston Undergraduate Academic Honesty Policy , University of Houston Graduate Academic Honesty Policy). I understand that academic honesty is taken very seriously and, in the cases of violations, penalties may include suspension or expulsion from the University of Houston."*

Helpful Information

Coogs Care: https://www.uh.edu/dsaes/coogscare/

Laptop Checkout Requests: https://www.uh.edu/infotech/about/planning/off-campus/index.php#do-you-need-a-laptop

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