COURSE NUMBER: HPE DSI 311
COURSE TITLE: Introduction to Machine Learning
INSTRUCTOR: Ioannis Konstantinidis
CONTACT: ikonstantinidis@uh.edu / 713-743-0669 / Online and by appointment

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 via MS Teams.

Course format: This course is being offered in the Synchronous Online format. Synchronous online class meetings will take place according to the class schedule. In between synchronous class meetings, there may also be asynchronous activities to complete (e.g., discussion forums and assignments). The final exam will be delivered in the synchronous online format.
Attendance: Attendance is mandatory. Grades (Pass/Fail) will not be assigned to students who fail to attend at least 12 hours of synchronous instruction.
Required Materials: None. Lecture notes will be provided via the course website.
Evaluation and Grading:

- 3 homework assignments: 22 points each (66 points total)
- Final exam: 34 points (last day of class)

Students must collect at least 70 points to pass the course.
Course description: At the end of this course, successful students will understand the logic behind and of the high-level structure of the algorithms implemented by common Python libraries (e.g., Pandas, NumPy, Matplotlib, Scikit-Learn) for EDA/ML tasks (e.g., estimators, preprocessing and scoring methods) and be able to present and discuss their process and findings by creating interactive documents using the Jupyter Notebook web environment for reproducible dissemination of EDA/ML results. Successful students will be able to perform

1. Exploratory Data Analysis (EDA) tasks using common Python libraries, including:
    a. Loading Data from flat CSV files containing structured, tabular datasets, including:
        i. local/remote file locations (e.g., local disk, OpenML.org)
        ii. variables of different types (e.g., categorical, numerical)
    b. Summarizing Data, including:
        i. numerical and graphical descriptions (e.g., tables, figures)
        ii. uni- and bi- variate descriptive statistics (e.g., centrality and dispersion measures, histograms, scatter plots)
    c. Preprocessing Data as appropriate, including:
        i. selecting specific columns/rows based on basic index or value criteria
        ii. transforming (e.g., scale, center)
        iii. factorizing (e.g., PCA for dimension reduction)
2. Supervised Machine Learning (ML) tasks using common Python libraries, including:
    a. Training models to predict classification/regression targets, including:
        i. basic predictive estimators (e.g., kNN, Logistic Regression, linear SVM, Decision Tree)
        ii. non-linear extensions such as kernels or bagging/boosting ensembles (e.g., Random Forest, Stochastic Gradient Boost)
        iii. multi-layer perceptron estimators with varying network architectures (e.g., # of layers / neurons per layer) and back-propagation optimization parameter combinations (e.g., solver, epoch, learning rate)

b. Selecting an optimally tuned model by assessing an individual estimators' performance potential, relying on:
    i. a common data and measurement framework (e.g., k-fold cross-validation, fixed randomization seed)
    ii. the exploration of several hyperparameter values (e.g., regularization)
    iii. considerations about model generalization (e.g., bias/variance trade-off within measured performance results)
c. Validating the selected model's performance using scoring methods appropriate for the problem/question, including: precision, recall, F-1 score, Cohen's Kappa, etc.

3. Unsupervised Machine Learning (ML) tasks using common Python libraries, including:
    a. Vector Quantization estimators (e.g., k-means) that partition the dataset into a fixed number of clusters
    b. Hierarchical Clustering estimators (e.g., agglomerative) that produce nested multi-level clusters (e.g., dendrograms) based on different linkages (e.g., single, complete)

Prerequisites: This course requires familiarity with some basic concepts which are present in most programming languages: primitive and non-primitive data structures and operators, conditional and repeated execution, and working with libraries/modules. All programming work in this course will be done using Python, a beginner-friendly language. Familiarity with "HPE DSI 212 scientific Programming with Python" or equivalent is required for this course.

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. 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 Center for Students with DisABILITIES. 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.

Academic Honesty: The University of Houston Academic Honesty Policy applies.

Counseling and Psychological Services: CAPS can help students who are having difficulties managing stress, adjusting to college, or feeling sad and hopeless. You can reach CAPS (www.uh.edu/caps) by calling 713-743-5454 during and after business hours for routine appointments or if you or someone you know is in crisis. Also, there is no appointment necessary for the "Let's Talk" program, which is a drop-in consultation service at convenient locations and hours around campus.