**Homework 4**                                                                **CSCE 633**
Due: 11.59pm on November 24, 2021

---

**Instructions for homework submission**
Please submit on eCampus a **single pdf** file containing your solutions.
a) Please write a brief report and **include your code right after each answer**.
b) For each answer, please explain your thought process, results, and observations. Please do not just include your code without justification.
c) Create a **single pdf** and submit it on **CANVAS**. Please do not submit .zip files or colab notebooks.
d) The maximum grade for this homework is **10 points** (out of 100 total for the class). There is **1 bonus point**.
e) **You can use any available library for this homework.**

**Question: Predicting one's hirability based on their job interview**

Communication skills are essential for successfully entering the workforce. A particularly important communication skill is the ability to connect the job role requirements with one's professional or personal experience, which is commonly asked during job interviews.

The goal of this problem is to predict one's job hirability based on their physiological responses and vocal measures during the interview. We have collected data from 13 participants during a job interview with an industry representative, located inside "Homework 4" folder on CANVAS ("data.csv") . The rows of the file refer to the samples of the dataset, where each interviewee participant corresponds to one sample. The columns denote the participant ID (column 1), participant's physiological and vocal measures (columns 2-9), and the outcome variable (column 10), as described bellow:

1. PID: interviewee ID

2. SCL: skin conductance level measured as a proxy of the amount of sweat elicited from the body (measure of physiological reactivity)

3. SCRAmp: average amplitude of skin conductance responses (measure of physiological reactivity)

4. SCRfreq: average number of skin conductance responses (measure of physiological reactivity)

5. HRmean: average heart rate (measure of physiological reactivity)

6. ACCmean: average wrist acceleration

7. Energy: energy of the speech signal (measure of voice loudness)

8. ZCR: speech zero-crossing rate (proxy measure of speech rate)

9. VoiceProb: voicing probability of speech (measure of speech quality)

10. Hirability: hirability score assigned by the interviewer (1-5 scale; 1: low; 5: high)

**(1) (2 points) Feature exploration:** Compute the Pearson's correlation between each feature and the hirability score (i.e., 8 correlations in total). Which features appear to be the most predictive of the outcome? Please comment on the sign (i.e., positive/negative) of the correlation values.

**(2) (3 points) Decision tree modeling:** Use a decision tree to estimate each interviewee's hirability score based on their physiological and vocal measures. Use a leave-one-sample-out cross-validation (i.e., equivalent to leave-one-participant-out in this case), according to which you will have as many folds as the number of samples (i.e., participants). In each fold, you will use one participant as the test and the rest for training the decision tree. Please collect the estimated hirability score on the test sample from each fold. After all folds are done, please report the average absolute error across all participants (i.e., by taking the estimated hirability value of the test participant from each fold). Experiment with various hyper-parameters (e.g., tree depth) and feature combinations.

**(3) (2 points) Decision tree visualization:** Provide a plot of the root, nodes, and decision boundaries of the best decision tree. Provide your intuition regarding the resulting model.

**(4) (3 points) Random forest:** Repeat the same task as in question (2) using a random forest. Experiment with the optimal tree depth and number of trees. Compare and contrast the performance of the decision tree with the random forest.

**(5) (1 point) Adaboost:** Repeat the same task as in question (2) using Adaboost with a base model of your choice. Compare and contrast the performance of Adaboost method with the decision tree with the random forest.