CSCI 5622-001
Homework 5: Designing ML models for a real-world problem

Instructions for homework timeline and submission

a) In this homework you will work within teams of 4 classmates. The team assignments are uploaded on CANVAS. You can reach out to your team members via CANVAS.
b) You will present your work with your team in class on **May 1st** during class time (**4.40-5.55pm MT**). Most of the work should be ready by the class presentations, including the data analysis and the e-poster. Participation in the e-poster presentation will count toward the in-class participation grade.
c) The final report is due on **May 5, 2024 @ 11.59pm**. Please create a zip file with two pdf files, including the final report and the e-poster. **One member per team can make the submission.**
d) You can use any publicly available library or code for this homework.
e) The total for this homework is **10 points** (out of 100 total for the class) and **1 bonus point**.

Speech-based ML technologies have observed an increased focus by digital healthcare due to the fact that speech can be unobtrusively collected via smartphones and wearable devices on a continuous basis, and carries valuable information about the human behavior and mental state. Speech is a result of the complex interplay between cognitive planning and articulation. The cognitive component involves cognitive planning via the formation of the message that a speaker intents to communicate. The motor component of speech production, also referred to as "articulation", relies on the coordination of the lungs, glottis (i.e., including the vocal cords), and vocal tract (i.e., mouth, nasal cavity). Both the motor and cognitive components of speech can be affected by the speaker's traits and states, the first reflecting one's permanent characteristics (e.g., race/ethnicity, gender) and the latter varying over time (e.g., mental health condition). Due to this richness of information, acoustic measures derived from speech, such as prosody or spectrotemporal characteristics, can reflect information that is critical for mental healthcare applications such as detecting depression, while at the same time they can be confounded by demographic factors. Hence, ML algorithms that rely on speech measures to detect mental health conditions can potentially treat individuals differently based on their demography, thus propagating existing disparities. This homework explores gender bias in speech-based ML algorithms that detect depression.

The data comes from the Distress Analysis Interview Corpus Wizard of Oz (DAIC-WoZ) dataset, which contains clinical interviews designed to support the diagnosis of psychological distress conditions such as depression. More details about the dataset, including the experimental setup and type of data, can be found here: `https://dcapswoz.ict.usc.edu/`.

The data uploaded on CANVAS includes 107 participants (63 male, 44 female). Out of the 107 participants, 30 participants have depression (14 male, 17 female). The 'features_train.zip' file includes 87 csv files (one per participant) with the training data, while the 'features_test.zip' file includes 20 csv files (one per participant) with the test data. Within both folders, the filename 'spk_x.csv', where x is the participant ID, includes the acoustic features from participant x. The rows of the file correspond to the turns of a participant during that interview and can vary in number depending on the interview length. The columns of the file correspond to 88 acoustic features that were extracted at the turn-level based on the Geneva Minimalistic Acoustic Pa-

rameter Set (eGeMAPS) and include parameters related to speech frequency, energy/amplitude, spectral balance, and timing. You can find more information in the 'feature_description.csv' and here: `https://ieeexplore.ieee.org/document/7160715`. The file 'labels.csv' contains three columns that correspond to the participant ID, the participant depression label (0: no depression, 1: depression), and participant gender (0: female, 1: male), including participants from both training and testing sets.

(a) (3 points) Classifying for depression and gender. Use a machine learning algorithm of your choice to conduct two types of classification tasks based on the original set of acoustic features.

(a.i) (1.5 points) Depression classification. In the first task, you will classify between speakers who have depression and speakers who do not have depression using the original set of 88 acoustic features. Please report the simple classification accuracy A and balanced classification accuracy BA on the test set for all participants included in the test set, as well as for the female and male participants in the test set separately. In addition, please report the equality of opportunity (EO) that computes the difference in true positive rate TPR (i.e., ratio of correctly classified participants with depression) between female and male participants, i.e., $EO = 1 - |TPR(male) - TPR(female)|$, quantifying to what extent the same proportion of female and male participants receive a true positive outcome. Please discuss your findings.

Note: Each turn can be used as a sample for the train and test data, thus, you will be obtaining a decision on whether an turn comes from a participant with depression or without depression. However, the accuracy and EO measures should be computed at the participant level. This means that you will need to aggregate the turn-based decisions on depression classification at the participant level (e.g., via averaging those decisions).

Note: The simple classification accuracy A and balanced classification accuracy BA are defined as follows:

$$A = \frac{\#\text{correctly classified samples}}{\text{total } \# \text{ samples}}$$

$$BA = 0.5 \cdot \frac{\#\text{correctly classified samples for depression}}{\text{total } \# \text{ samples for depression}} + 0.5 \cdot \frac{\#\text{correctly classified samples for no depression}}{\text{total } \# \text{ samples for no depression}}$$

(a.ii) (1.5 points) Gender classification. The second is a gender classification ask, where you will classify between female and male speakers. Report the simple classification accuracy and balanced classification accuracy. Please discuss your findings.

Note: Each turn can be used as a sample for the train and test data, thus, you will be obtaining a decision on whether an turn comes from a female or male participant. However, the accuracies should be computed at the participant level. This means that you will need to aggregate the turn-based decisions on gender classification at the participant level (e.g., via averaging those decisions).

(b) (1.5 point) Finding the most informative features for depression. Explore a filter feature selection method of your choice to identify the n features that are the most informative of depression on the training data. Report and discuss the 20 features that are most indicative of depression. Use the same machine learning algorithm as in question **(a.i)** to classify between depression and no depression based on n number of selected features, where $n = 10, 15, 20, 25, 30, \ldots, 50$. Plot the simple and balanced depression classification accuracy

(for all participants, as well as female and male participants separately) and EO on the test set against the number n of selected features. Please discuss the results.

Note: Similarly to question (a), classification accuracies and EO should be reported at the participant-level. The simple classification accuracy A and balanced classification accuracy BA are defined as follows:

$$A = \frac{\#\text{correctly classified samples}}{\text{total } \#\text{ samples}}$$

$$BA = 0.5 \cdot \frac{\#\text{correctly classified samples for depression}}{\text{total } \#\text{ samples for depression}} + 0.5 \cdot \frac{\#\text{correctly classified samples for no depression}}{\text{total } \#\text{ samples for no depression}}$$

(c) (1.5 point) Finding the most informative features for gender. Explore a filter feature selection method of your choice to identify the m features that are the most informative of gender on the training data. Report and discuss the 20 features that are most indicative of gender. What do you observe in relation to question **(b)**? Use the same machine learning algorithm as in question **(a.ii)** to classify between female and male speakers based on the m number of selected features, where $m = 5, 10, 15, \ldots, 30$. Plot the simple and balanced gender classification accuracy against the number m of selected features on the test data. Please discuss the results.

Note: Similarly to question (b), classification accuracies should be reported at the participant-level.

(d) (2 points) Mitigating bias via removing gender-dependent features. Remove the m most informative features of gender from the original feature set. Using the revised feature set, use the same machine learning algorithm as in **(a.i)** to conduct depression classification. Please report the results similar to **(a.i)** and discuss your findings.

(e) (1 point, Bonus) Mitigating bias via other approaches. Experiment with other bias mitigation approaches such as the ones described in class. For example, you remove potential bias via re-weighting the samples between the female and male speakers when training the machine learning model, or via applying adversarial learning for reducing the evidence of sensitive attributes in the data. Please report the results similar to **(a.i)** and discuss your findings.

Note: You can take a look at the following papers for more ideas:
`https://dl.acm.org/doi/10.1145/3278721.3278779`
`https://dl.acm.org/doi/fullHtml/10.1145/3462244.3479897`
`https://www.sciencedirect.com/science/article/pii/S2772442522000430`

(f) (1 point) E-poster. Create an e-poster presentation of your work. The e-poster will provide the main gist of your work, including the problem statement, your methodology, and the main results from your experiments. **Add visuals to your poster so that people understand the main concepts.** Do not make your e-poster too crowded, since you want other people to be able to see through the screen projection. You can find here the link to prepare your poster presentation `https://www.youtube.com/watch?v=1RwJbhkCA58&feature=youtu.be`.

Note: Each team will present their poster via a laptop in class on **May 1st**. Half of the teams will be presenting in the first half of the class, while the other teams will go around the posters and discuss their classmates' solutions. Roles will switch during the other half of the class.

(g) (1 point) Teamwork. Please follow the survey link below, in which you will need to discuss the level of engagement of your team members. The feedback that your team members will provide for one another will be used as an indication of each member's contribution to this group assignment.

Survey Link: `https://forms.gle/GxNqFBfxXtJbPqVU6`