

Hangyul Yun, Gabbriel Bui, Angel Solorzano, Martin Nguyen, Sofia Cho

ECS 171

Professor Rafatirad

Mid-Quarter Progress Report

Done: Created color-coded scatter plots to aid in finding patterns in data, created correlation matrix heatmap, created separate plots for each antenna

Todo: Visualize the data of more patients, determine if it is possible/useful to combine the data of multiple patients, create and train models

Based on the generated correlation matrix heatmap alongside the scatter plots, we have deliberated to use frontal, vertical, and lateral acceleration as well as frequency and time in order to predict activity. However, there is still a lot of noise in the data. We will need to normalize it first in order to finish the Data Preprocessing step. From here, we can take our next steps.

Currently, our findings come from data compiled from one specific data file, as a cursory analysis. In order to better understand our dataset, we can generate correlation matrix heatmaps from different files included in this dataset and see how the relationships between the features vary across more of the whole dataset, as opposed to exploring how the features relate as it appears in just one datafile.

In consideration for the project's completion and the final deliverable, we currently have plans to use an Artificial Neural Network in order to model our predictions. The linear regression models that were learned in the beginning of class were largely deemed unsuitable based on our cursory data analysis revealing that the data did not appear linear. Thus, using an Artificial Neural Network in order to create our predictions is our current plan. However, as more of the quarter progresses and as we learn more about other Machine Learning models, this could be subject to change if a different model turns out to be better suited for the task.

As of now this is what our current roadmap looks like.

Finish preprocessing data - May 10, 2024

Finish exploratory data analysis - May 17, 2024

- ✿ Find a way to interpret frontal, vertical and lateral acceleration and how it affects our dependent variable
- ✿ Determine whether gender significantly influences the data, enough to include gender in our final report (reach feature?)

Finish algorithm - May 31, 2024

- ✿ Algorithm should work on one datafile
- ✿ Implement main algorithm
- ✿ Implement model evaluation
- ✿ Modify code to iterate through all files

Finish report - June 8, 2024

- ✿ Section out parts to each other

As we reach each part of the roadmap, we will update it with mini milestones and also update any dates as well.

Currently, the greatest obstacle currently facing the progression of the project is understanding the relationship between the three acceleration features (vertical, frontal, and lateral), what they may mean (in case we favor one of these features more strongly than the others, or deem one of them less crucial to creating a comprehensive Machine Learning model), and if they end up influencing other features that contribute to the final activity values.