My goal for this project (besides passing the course) is to add to my professional portfolio that is targeted toward the “Health Sensing ML Engineer” position at Apple, which works on “developing algorithms for a variety of health sensors, including PPG, accelerometer, and ECG.” The minimum qualifications for this position also include “Sufficient SW skills to run large ML training jobs efficiently on a distributed backend with large volume of data”, which I intend to demonstrate with distributed training of my neural regression models on Google Colab, even if my dataset isn’t large enough to realistically require it (realistically I could probably get away with regressing using something simpler like XGBoost, but it’ll be important for me to demonstrate my abilities with PyTorch).

The goal of the project itself will be to use accelerometer and raw heart sensor data to estimate calorie expenditure from activity. The best dataset I’ve found so far to accomplish this is the PAMAP2 Physical Activity Monitoring Dataset, which includes both accelerometer and heart rate data from 9 subjects performing a variety of activities. These activities were controlled for MET (metabolic equivalent of task) values at the time of measurement, which I can use as the gold label for training a regression model. These MET values can then be converted into calories (using well-established formulae) to produce my final estimation. These data are quite raw from what I understand, which should provide a good opportunity to demonstrate preprocessing and feature engineering.

As a bonus if I have extra time, I think it might be cool to also take some well-established step-counting algorithm (which I’m hoping I could find in the literature) and apply it to my accelerometer data so that I can also analyze the relationship between these steps and my calorie estimation. This analysis could support or refute the value of providing a step-count to users as a metric to guide their physical activity, which is becoming more common in the sports-science/fitness community, even at more advanced levels (Whoop for example, just recently added a step-tracking feature to their app).

What do you think about this project? Is there anything I might add/take away to better serve the purpose of this class and the purpose of bolstering my professional portfolio? What sort of details should I pay attention to that I might otherwise overlook? Thanks.