Mica Haney
Sarah Beaver

Progress Report 1

- What progress has been made so far, and how it lines up with the plan and timeline submitted

We have finished creating an account on allofus. Mica was not able to verify identity during the account creation process in the first week. It took another half a week to get the verification figured out for her account. Both account has proof of identity completed, finished the allofus training, and passed all the tests necessary to access the data on allofus.

A cohort was created that included patients that have taken the survey or have conditions. We selected the data to include all available columns that we could use in each of the following tables (conditions, surveys, and patient demographics). Then, we created a workspace where we can write python code. We changed which conditions we are looking into to be acute myocardial infarction (heart attack), rheumatoid arthritis, diabetes, and sleep apnea. We selected patients that had a icd 10 or icd 9 code for one of the conditions listed previously.

We were able to merge the tables together so that it is one patient per row and to have the questions assigned to specific columns instead of randomly merging together. We dropped columns that we did not think were relevant such as code ids, some source name, source values, ect.

- What's been done since the last progress report?

This is the first progress report so there has not been a previous progress report.

- What challenges or obstacles have arisen

The first obstacle was creating accounts and verifying identity for Mica, which merely took time to finish. The second obstacle faced was merging the data and transposing it to be one patient per row. Which we continued into this last week. Another obstacle has been that the environment being used does not support two users working on the same notebook simultaneously well. The environment forces one person to abandon their work and reload periodically. This has been largely avoided by both of us working at different times from each other.

- What the next steps are – what do you expect to complete between now and the next report?

Mica Haney
Sarah Beaver

We need to finish some data selecting and training/testing splitting. We expect to have two models created and finished with only needing to work on hyperparameters. We will also have started writing on the final paper.