

## **Sprint 4 Plan**

Understanding Healthcare Data  
Sprint Completion Date: 5/24/2020  
Revision 1.1, Date: 5/11/2020

Goal: Expand the current dataset and dockerize it so it may be run on real data, while also beginning work on the end of quarter presentation.

### User Stories

**User Story 1:** *As a company, I want to run the models on a set of real patient data to test its effectiveness in a real world scenario.*

**Task 1:** Create a more effective dataset in which to train our model. (8 points)

- Vary the window size in order to see if that allows for better training
- Use Synthea to generate 50k more patients without filtering for a disease to use as 0 labels for congestive heart failure and myocardial infarction.

**Task 2:** Analyze the data to determine the source of our high accuracy. (7 points)

- Evaluate the way our data makes determinations
- Check the data to ensure we're receiving the data we're interested in during every stage of the extraction process.

**Task 3:** Dockerize the components so they can be run on real data. (5 points)

- Fine tune the docker container to fit the newly created datasets
- Transfer the docker containers to Rob so he can run them on the actual healthcare data from Anthem.

Total for user story: (20 story points)

**User Story 2:** *As a group, I want to create an organized and informative presentation to demonstrate the work completed over the duration of the project.*

**Task 4:** Develop a powerpoint presentation that encompasses our project. (10 points)

- Create slides and graphics that convey the meaning of our project, goals, timelines and work done.

Total for user story: (10 story points)

### Initial Task Assignment:

Cassidy Norfleet - task 1, 2, 4

Brendan Reilly-Langer task 1, 2, 4

Aman Prasad - task 2, 3, 4

Harshitha Arul Murugan - task 2, 3, 4