Sprint 3 Plan

Understanding Healthcare Data

Sprint Completion Date: 5/10/2020

Revision 1.1, Date: 4/26/2020

Goal: Test the CNN and RNN using data from a different disease (Myocardial Infarction) to ensure that our results are accurate and start preparing the models to be sent to Anthem so they could be tested on real data.

User Stories

User Story 1: *As a programmer, we want to know if the models are performing accurately.*

Task 1: Ask Rob to generate Myocardial Infarction data and send it to us. (2 points)

Task 2: Neatly package the data so it can be run on the models. (5 points)

Task 3: Edit the models to fit the data. (5 points)

Total for user story: (12 story points)

User Story 2: *As a programmer, I want to receive model-ready data right after feeding raw data into a single script.*

Task 4: Create windows of embeddings for each patient (10 points)

- Generate 6 month windows of lists of strings
- In parallel, collapse the windows into one large list of strings

Task 5: Train a Word2Vec model on the sentences (5 points)

Task 6: Using the data in the Word2Vec models, generate embedding matrices and dump them into NPY files (5 points)

Total for user story: (20 story points)

User Story 3: *As a programmer, we want to be able to send the models to Anthem*

Task 7: Turn training notebook into a Python script (5 points)

- Download and refactor code such that it will read in a .npy file and perform the necessary transformations to feed the data into NNs

Task 8: Prepare a docker container to run all of the training scripts. (3 points)

- Should take in CSVs and NPY files

Total for user story: (8 story points)

User Story 4: *As a group, I want to create an organized poster to demonstrate the work completed over the duration of the project.*

- **Task 9:** *Develop a poster that summarizes our project. (10 points)*

Total for user story: (10 story points)

Initial Task Assignment:
Cassidy Norfleet - task 1, 2, 4, 5, 9
Brendan Reilly-Langer task 1, 2, 4, 5, 9
Aman Prasad - task 3, 7, 8, 9
Harshitha Arul Murugan - task 3, 7, 8, 9