

Sprint 1 Plan

Understanding Healthcare Data
Sprint Completion Date: 4/12/2020
Revision 1.2, Date: 3/29/2020

Goal: Create a model to train on word embeddings created given the English descriptions of medical codes present in a patient's data in a given month.

User Stories

User Story 1: *As a programmer, I want to be able to convert the English descriptions of medical codes present in a patient's data in a given month*

Task 1: Generate "sentences" of patient data given the different time buckets through a multithreaded process (10 points)

Task 2: Train a Word2Vec model (5 points)

Task 3: Generating matrices with the Word2Vec model, 1 vector per sentence (5 points)

User Story 2: *As a healthcare provider, I would like to be able to run the data on multiple models to compare and ensure that the predictions are correct.*

Task 4: Modify the CNN to use the embedding matrices (5 points)

Task 5: Modify the RNN to use the embedding matrices (5 points)

Initial Task Assignment:

Cassidy Norfleet - task 1

Brendan Reilly-Langer task 1

Aman Prasad - task 3, 4, 5

Harshitha Arul Murugan - task 3, 4, 5

**Had a lot of spikes as we had to learn how to use the word2vec model*