Product name: Understanding Healthcare Data
Team name: Anthem Sponsored
Release name: V3.0
Release date: 3/13/2020
Revision number: 3.0
Revision date: 5/10/2020

High-Level Goals:

- Can you predict the occurrence of a given chronic disease given an arbitrarily sized window of medical history?

User Stories:

- Sprint 1 (3/30/2020 - 4/12/2020)
    - 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.

- Sprint 2 (4/13/2020 - 4/26/2020)
    - Get running on hummingbird
        - Figure out if we need access to AWS accounts
    - Determine if the model can be generalized to myocardial infarction
    - Refactor data pre-processing/transformation notebook so additional conditions can be added.
    - Work with larger data set to train model
    - Start specifying how to package up model and data pipeline so Rob can run it
        - Build some trivial container as a start

- Sprint 3 (4/27/2020 - 5/10/2020)
    - Start making poster for the project
    - Apply model to a chronic disease: myocardial infarction
    - Reassess which codes should be included or discluded. (Filter out common/generalized codes)
    - Sum the codes for chf and myocardial infarction to determine which codes are contributing the most towards making correct predictions.
    - Combine scripts to generate sentences and embedding matrices
    - Turn training notebook into a Python script
    - Prepare a docker container to run all of the training scripts, taking in CSVs and NPY files

- Sprint 4 (5/11/2020 - 5/24/2020)
    - Have Rob run on real data
    - Fine tune the docker container to fit the new datasets
    - 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.
    - Create the powerpoint presentation

- Sprint 5 (5/25/2020 - 6/7/2020)
    - Generate and train on a dataset with the oldest 12 months zeroed out to compare the effectiveness of our RNN vs. our CNN.
    - Add the RNN to the Docker container.