DRG Determination via ML Methods

* What is the problem are you trying to solve? This must be related to the biomedical domain of knowledge.

Diagnosis-Related Groups (DRG) are a way to group patients based on similar diagnoses, procedures, and other factors to standardize the cost needed (as there is a fixed cost associated with each DRG and it’s treatment) for treatment for these groups. Determining DRG’s via ML methods given clinical notes and data can be very beneficial to many parties - providers, payers, etc. - in not only determining DRGs, but validating charts that have been coded with a particular DRG. Automating, or mostly automating this process via ML methods can save time and resources in DRG determinations.

* What are your main objectives for this project? Why is this problem important to solve using machine learning?

DRG determination/validation can be done by humans. However, with the volume and velocity of healthcare data, it’s essential that processes be automated. I believe this can be done through Machine Learning. Especially when there is a vast amount of unstructured free notes, a source of information that often goes largely unused, in clinical charts.

* What dataset(s) will you be using?

MIMIC-III is a dataset that includes data from more than 40,000 patients admitted to the ICU at Beth Israel Deaconess Medical Center. The data contained spans a range of more than a decade and includes a variety of data types: free notes, test results, medications, diagnoses and so on. And perhaps, most importantly, the data has been de-identified and is open to the public. This dataset should be sufficient to serve as a basis for developing a ML model in predicting DRGs in charts because one: the dataset’s extensive size, and two: since this is real data, it should be able to be applied to other datasets and real world data because it is real data.