Analytics Approach

Doctor writes the notes on the computer and later makes a bill by searching through a lot od ICDs and CPTs. There are around 20000 ICDs and 13000 CPTs.

Data Requirements

1. Historical notes of the doctors regarding different patients
2. The ICDs that were applied against those notes
3. The CPTs that were applied against those notes.
4. The billing information

Data Collection

The data will be collected through different sources.

1. The billing department, nurses and the doctor notes and placed in a single database.
2. CPTs will be gotten from https://www.medicalbillingandcoding.org/intro-to-cpt/
3. ICDs will be crawled from https://www.icd10data.com

Data Understanding and Preparation

1. Data will be joined together so one could know which CPT was with which note and so for the ICD
2. ICDS will be mapped to the single version
3. Any value that contains NULL will be removed.
4. It will be cleaned using various techniques
5. The Note, ICD and CPT will be the features to work on

Modeling and Evaluation

Different models will be applied to get the results.

1. Sequential Models like RNNs and LSTMs will be applied giving the result of CPT and ICD next to it
2. It will be a multi-label problem and model will be built keeping that in mind
3. Hyper parameters will be tuned using various ML techniques
4. An ensemble of different techniques will be able to give us the required results

Evaluation will be done on the following parameters

1. The f-score for both ICDs and CPTs will be calculated
2. Recall and Precision will be required.
3. A high level of accuracy will be required.

This will help in reducing the time required for creating a bill