**Write-up-** I have used **Markov Chains** for prediction of next events of Diagnosis. I have used only the ICD-9 data for training(creating a occurance\_dictionary). Using markov chain was inspired by the idea that all if there is a pattern that a particular disease occurs right after one, this can be well captured by Markov chain which depend previous states.

**Quality Checks/Error Handling**- If there is a disease no.(ICD-9) which occured for the 1st time for any patients as their last 2 datapoints(in the time-series), as there is no record of it in the occurance\_dictionary, hence it won’t be able to predict for that particular patient. For avoiding this I have taken a reserve of past values to be used as the reference key for prediction.

E.g. If **2386, 4563, 6553, 2377** are the last 4 ICD-9 coded disease occurred to the patient, and the **2377** disease occurred for the **1st** time in the data, my model will take **6553** as reference and predict from that point. And if this also occured for the first time(which is highly unlikely), it will take **4563** as reference.

**Data preprocessing steps:**  In the preprocessing stage all the ICD-10 values were removed taking only numerical values for training.

**Expected Error :** The accuracy will be around 9-15% as in case of the public data.