

Data Description:

file format:

train.parquet

test.parquet

sample_submission.csv

final_submission.csv

Health records of patients diagnosed with a certain disease are present in both train and test

files, various categories of health records like diseases diagnosis, tests, symptoms, and drug treatments are captured in the records along with the date of occurrence for each patient

Patient-Uid: Unique Identifier for each patient

Date: Date on which the patient encountered the event

Incident: Actual event encountered

Total Number of unique patients present in train.parquet 27K

An event called "Target Drug " is present in the incident column of the training set for around 9K

patients at least once in their journey.

Problem Statement:

Drugs are developed in therapeutic areas to boost the patients' condition against chronic and terminally ill diseases, the "Target Drug" is one such and it can boost the patients' health without making them dependent on the other drugs that can lead to life-threatening side effects. The objective is to build a predictive model to estimate if a patient is eligible for the first prescription of the "Target Drug" in the next 30 days so that the physician who is treating the patient could be informed on the better treatment choices.

A. Come up with the right strategies to create positive and negative data samples from the data(avoid any biases while sampling the data to build a good predictive model)

B. Build a predictive model by doing the appropriate feature engineering(eg: frequency-based, time-based features, e.t.c). The predictive model could also leverage Deep Learning based techniques.

C. Evaluate the model on your own validation set and come up with the right threshold to minimize false positives and false negatives

D. Some of the patients present in the test file are eligible for the drug prescription within a month and some of them are not, using each patient's historical data predict if he/she is eligible for the "Target Drug"

E. Each patient-uid should be labelled with a binary value of 1 or 0 using the built model, 1 is considered as eligible for the "Target Drug" in the next 30 days and 0 considered as un-eligible