Length of Hospital Stay Prediction

DAB 402 – Capstone Project

Assessment 3 – Detailed data assessment

To: Mr. Mohammad Shahid

From: Group 11

Jaspreet Kaur	0730470
Kanchan Bagga	0732356
Varinderjit Singh	0730482

Date – 16 Feb, 2020

Problem Statement:

The ability to predict the length of stay (LOS) as early as possible in the preadmission stage after first checkup might be helpful to monitor quality care for the hospital admission management team. In this project we shall develop model to predict the LOS at admission time for general patients.

Dataset:

In our dataset we have 5,60,486 observations and 972 columns. We have enough data for training and testing purpose to predict accuracy. This is uncleaned dataset. We have many of NULL values in our dataset. We may need to drop some or many unnecessary variables and observation for the cleaning purpose.

Dataset source:

We have downloaded dataset github website. Reference is given below:

https://github.com/yaleemmlc/admissionprediction/blob/master/Results/5v_cleandf.RData

Five C's:

i) Consent:

As this dataset is opensource, we don't need to have any consent to collect this dataset.

ii) Consistency:

We have huge volume of records in our dataset. It is very reasonable for experimenting on this dataset. Moreover, the accuracy of different models will be comparable and consistent.

iii) Clarity:

This data is used for model building with highest accuracy to predict length of stay (LOS) for general patients in hospital. So, It is very clear that how we use this data.

iv) Control:

Although, this dataset is publicly available, but it is controlled by the GitHub. Now we have access to this data, but we can use it only for the experimenting purpose. We cannot manipulate it for public but can make some changes (cleaning) according to our requirements only.

v) Consequences:

This data collection can never harm any individual. Instead, it will help in the better caring in hospital by predicting length of stay. The experiments on this dataset are going to help a lot to hospital admission management team.

Challenges:

We faced too many problems in getting dataset. After struggling many days we got this dataset anyhow but due to its huge volume and file type we again faced the problem of opening and changing its type from RData to CSV. We are still facing the problem of opening it in excel.

References:

https://www.hindawi.com/journals/jhe/2016/7035463/

https://github.com/YaronBlinder/MIMIC-III_readmission