Final Report

On

Length of Stay Prediction

Varinderjit Singh, Kanchan Bagga, Jaspreet Kaur

**Abstract:**

Vulnerability long of patient’s medical clinic stay is a significant obstruction to viable planning for confirmation of elective patients. Different techniques are introduced for diminishing this vulnerability, including of demonstrative data, doctors’ evaluations and medical caretakers release expectation. Neural systems give another strategy to assessing injury quiet level of disease and precisely foreseeing a patients’ length of remain at the basic consideration office. Backpropagation, outspread premise work is actualized to determine the relevance of neural systems for anticipating either injury seriousness or length of stay. Neural systems perform well on this clinical space issue.

**Keywords:**

Length of Stay, Admission stage, Machine Learning, SQL.

**Introduction:**

Predicting emergency clinic length of stay is of significant incentive for medical clinic asset arranging and the board. Having the option to more likely gauge how much longer a patient will stay in medical clinic, can help with booking the use of wards and emergency clinic beds including the planning of elective medical procedures dependent on forthcoming accessibility of beds. For emergency clinics it is alluring to enhance the utilization of beds to best give care, for clinicians prescient model scan give adjunctive clinical choice help, for patients improved arranging and expectation can add to their quality of care and for payers they are persistently seeking tools to build cost examination and forecast. Likewise, them towards esteem-based consideration requires more prominent expectation and optimization of how to keep up the well-being of a population. Being ready to more readily foresee length of remain and hence forth better care for patients is a significant piece of significant worth-based consideration.

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:**

For this project we use MIMIC-III database. MIMIC-III integrates deidentified, comprehensive clinical data of patients admitted to Beth Israel Deaconess Medical Center in Boston, Massachusetts, and makes it widely accessible to researcher internationality under a data use agreement. The open nature of data allows clinical studies to be reproduced and improved in ways that would not otherwise be possible. In this dataset provided 26 tables for different prospective.

**Literature Review:**

Article 1:

Predicting Hospital Length of Stay Using Neural Networks on MIMIC III Data

In this paper the authors are exploring the study of prediction for length of stay (LOS) for general patients using the MIMIC III database. They have used three different models those are: Support Vector Machine (SVM), Neural Network, and Decision Tree. They have trained a neural network to predict the patient’s stay in hospital in the number of days (how many days a patient will stay). They mentioned that out of three models SVM was the most accurate. Their prediction accuracy is approximately 80% and using linear model. They have mentioned that their database contains more than 50,000 records of people admitted to ICU units for 12 years (2001 to 2012). They have used 28 variables from the dataset.

Reference:

https://www.researchgate.net/publication/324177552\_Predicting\_Hospital\_Length\_of\_St ay\_Using\_Neural\_Networks\_on\_MIMIC\_III\_Data

Article 2:

Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network

This article shows the use of the neural network techniques to predict the Length of Stay for patients in a cardiovascular unit with one of three primary diagnoses: heart failure (HF), acute myocardial infarction (AMI), and coronary atherosclerosis (CAS). They have mentioned the variations in length of stay on two factors. One factor is hospital characteristics and other one is patient characteristics. They have explored the data for the National Health Service (NHS) in the United Kingdom. In this article they have mentioned that they have collected total 2,424 admission cases for three diagnoses. 872 heart failure (HF) patients, 572 acute myocardial infarction (AMI) patients, and CAS (coronary atherosclerosis) 933 patients. All these patients are over 65 years. They have analyzed the data from October 1, 2010, and December 31, 2011. Artificial Neural Network (ANN) is used in specific areas, such as cervical cytology and early detection of acute myocardial infarction (AMI). This article shows ANNs are more useful in predicting medical outcomes as compare to logistic regression, due to their nature of nonlinear statistical principles. Using ANN or linear regression model was able to predict correctly for 88.07% to 89.95% CAS patients at the pre discharge stage and for 88.31% to 91.53% at the preadmission stage. For AMI or HF patients, the accuracy ranged from 64.12% to 66.78% at the pre discharge stage and 63.69% to 67.47% at the preadmission stage.

Reference:

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

Article 3:

Predicting hospital admission at emergency department triage using machine learning

This study shows that machine learning can robustly predict hospital admission at emergency department (ED) triage and that the addition of patient history improves predictive performance significantly compared to using triage information alone. In this study it is mentioned that the data was collected for all adult emergency department (ED) visits from March 2014 to July 2017. They have used 972 variables to record each patient visit history. They have used three techniques (Gradient boosting, logistic regression, and deep neural network) for prediction. They have used three dataset types: one for triage information only, second for patient history only, and third for full set of variables. A total of 560,486 patient visits were included in the study. They have used patient's disposition as primary response variable that is encoded in a binary variable (1 = admission, 0 = discharge).

Reference: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201016>

Article 4:

Predicting hospital length-of-stay at time of admission

This article shows US hospital stays cost at least $377.5 billion per year on health system. It is mentioned that if we have prior knowledge of LOS, it can help in room and bed allocation planning also. They have used MIMIC dataset to implement the prediction model. they split the LOS target variable and features into training and testing data sets using the ratio of 80 and 20 respectively. Using the training set, they have fit five different regression models and then compared the accuracy on the testing dataset. They have compared the accuracy of GradientBoostingRegressor model and RandomForestRegressor model. As compare RandomForestRegressor they got more accuracy with GradientBoostingRegressor on testing dataset. The gradient boosting model RMSE is better by more than 24% (percent difference) versus the constant average or median models. They have used the information such as subject id, hospital admission id, admission date/time, discharge time, and many more. In their dataset they have 58,976 admission events and 46,520 unique patients They have used LOS in days as their target variable.

Reference:

https://towardsdatascience.com/predicting-hospital-length-of-stay-at-time-of-admission55dfdfe69598

Article 5:

Analysis of length of hospital stay using electronic health records: A statistical and data mining approach

In this article they have mentioned that they have used the database of patients admitted to a tertiary general university hospital in South Korea between January and December 2013. They have analyzed the patients according to the three categories. Those categories are descriptive and exploratory analysis, process pattern analysis using process mining techniques, and statistical analysis and prediction of LOS. Right now, EHR information and procedure mining innovation were utilized to break down all occasion logs entered among confirmation and release of the patient. This study helps to find the key factors correlating with duration of hospital stay at the prediction stage. The point of this investigation was to decide a strategy that could be applied to assist medical clinics in dealing with the length of inpatient remain more proficiently. In the data preparation phase, they extracted the EHR log data. Then data cleaning process was performed to extract meaningful analysis results. In the data analysis phase, they have used four types of analysis: LOS performance analysis, LOS analysis of transfer patterns, LOS analysis according to diagnosis, and analysis of long-term hospitalization. At the prediction phase, they identified the main factors correlating with the number of days of stay through data analysis and log-based statistical analysis.

Reference:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5898738/