LITERATURE SURVEY

Date	20 September 2022
Team ID	PNT2022TMID09355
Project Name	Analytics For Hospitals' Health-Care Data

1.Robust Length of Stay Prediction Model for Indoor Patients [Ayesh Siddiqa, Syed Abbas Zilqurnain Naqvi et al,2021]

In this paper a robust model is designed to help hospital administration to predict patients' Length Of Stay (LOS) to resolve the issues faced by hospitals.They collected a very large-sized data (more than 2.3 million patients' data) related to New-York Hospitals patients and containing information about a wide range of diseases including Bone-Marrow, Tuberculosis , Intestinal Transplant, Mental illness, Leukaemia, Spinal cord injury, Trauma, Rehabilitation, Kidney and Alcoholic Patients, HIV Patients, Malignant Breast disorder, Asthma, Respiratory distress syndrome, etc. have been analyzed to predict the LOS .The main objectives were to explore the Inpatient De-identied data and tobuild a robust model that could predict the hospital LOS of patients coming to the hospital in future. The conclude that the maximum stay was between 0 to 5 days with the meantime of each patient 5.3 days and more than 50 years oldpatients spent more days in the hospital. Visualization Exploratory data analysis (EDA) was used to analyze the dataset and summarize the dataset's main variables.

2. Predicting Length of Stay Across Hospital Departments[Miguel-Angel Sicilia, Elena García-Barriocanal et al,2021]

In this work they report a predictive model for length of stay (LOS) together with a study of trends and patterns that support a better understanding on how LOS varies across different hospital departments and specialties. Healthcare systems generate large amounts of administrative data about patients, departments, medical material costs, bed availability, diseases, etc. This departs from readily available administrative data toassess resource use in

hospital systems. Machine learning techniques are applied to hospital management in an attempt to optimise hospital resources more efficiently within the departments, providing an extra advantage in favour of patients and hospital entities , material and methods , data preparation, training and testing data setup ,evaluation criteria of the methods.Hospital departments were compared according to the algorithms used to predict results, taking into account those which offer better results in comparative tests.This research shows that reductions in hospital costs improvement in quality patient care are possible.

3. Risk factors predicting hospital length of stay in older patients with type 2 diabetes with Covid-19[Bilal Katipoglu ,Mehmet Ilkin Naharci et al, 2022]

In this study type 2 diabetes(T2D) patients with moderate Covid-19 symptoms were involved. Clinical treatments, laboratory and demographics data were extracted from the medical records. Univariate and multivariate logistic regression analyses were used for exploring the risk factors associated with Length of stay. They included 87 hospitalized T2D patients with mild or moderate Covid-19. Prolonged LOS was defined as equal to or greater than the median hospitable stay time. They conducted a secondary analysis using data from a prospective cohort study of patients aged 60 years and older hospitalized with non-severe Covid-19 pneumonia between August 1 and November 31, 2020 in a pandemic hospital in Turkiye. The endpoint of this study was hospital length of stay (LOS), which was calculated according to the number of days of hospitalization. Patients were divided into two groups according to the median LOS value: ≤ 7 days as normal and >7 days as prolonged LOS. In addition, this study reinforces the need for further research in larger cohorts to increase our knowledge of the factors afecting the hospital stay of older Adults due to Covid-19, especially given the high risk of hospital complications in those with diabetes.

4. The impact of inpatient bed capacity on length of stay[Brendan Walsh, Samantha Smith et al,2021]

Large reductions in inpatient length of stay and inpatient bed supply have

occurred across health systems in recent years. However, the direction of causation between length of stay and bed supply is often overlooked. This study examines the impact of changes to inpatient bed supply, as a result of recession-induced healthcare expenditure changes, on emergency inpatient length of stay in Ireland between 2010 and 2015. U- shaped trends are observed for both average length of stay and inpatient bed supply between 2010 and 2015. A consistently large positive relationship is found between bed supply and length of stay across all regression analyses. Between 2010 and 2012 while length of stay fell by 6.4%, our analyses estimate that approximately 42% (2.7% points) of this reduction was associated with declines in bed supply. Changes in emergency inpatient length of stay in Ireland between 2010 and 2015 were closely related to changes in bed supply during those years. The use of length of stay as an efciency measure should be understood in the contextual basis of other health system changes. Lower length of stay may be indicative of the lack of resources or available bed supply as opposed to reduced demand for care or the shifting of care to other settings.

5. Length-of-Stay Prediction for Pediatric Patients With Respiratory Diseases Using Decision Tree Methods[Fei Ma, Limin Yu et al,2020]

In this paper we are going to see about the prediction of LOS for pediatric patients with respiratory diseases. The are predicting the LOS by using the Decision Tree Methods. We use three types of decision tree they are bagging ,adaboost and random forest. There are two test which are used to assist the performance of prediction methods they are bisection test and periodic test. All the three methods are effective but adaboost is slightly better than other two. Bootstrap aggregation (Bagging) was proposed by Leo Breiman to improve classification accuracy. Adaptive boosting algorithm (AdaBoost) represents a popular ensemble method that uses simple classifiers, called weak learners, to generate a stronger classifier. Random forest is also a type of ensemble learning classifier that involves independently growing decision trees based on independent training set. In bisection test bagging method performs the best and in periodic test adaboost performs the best .By predicting the LOS of patients accurately we can create more efficient and effective

management of health care resource.

6. Predicting Length of Stay for Cardiovascular Hospitalizations in the Intensive Care Unit: Machine Learning Approach[Belal Alsinglawi, Fady Alnajjar et al,2020]

In this paper we are going to see about the prediction of LOS for cardiovascular hospitalizations in the intensive care unit. The are predicting the LOS by using the machine learning approach. They are going to predict the LOS for patients admitted in coronary care unit(CCU) and in cardiac intensive care units(CICU). This paper introduces a predictive research architecture to predict Length of Stay (LOS) for heart failure diagnoses from electronic medical records using the state-of-art- machine learning models in particular, the ensembles regressors and deep learning regression models. Our results showed that the Gradient Boosting Regressor (GBR) outweighed the other proposed models in this study. The GBR reported higher R-squared value followed by the proposed method in this study called Staking Regressor. Additionally, The Random forest Regressor (RFR) was the fastest model to train. Our outcomes suggested that deep learning-based regressor did not achieve better results than the traditional regression model in this study. This work contributes to the field of predictive modelling for electronic medical records for hospital management systems.

Project Description

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas to focus: Healthcare

Management. While healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital.

This parameter helps hospitals to identify patients of high LOS-risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning.

Suppose you have been hired as Data Scientist of Health Man – a not for profit organization dedicated to manage the functioning of Hospitals in a professional and optimal manner.

Goal:

The goal is to accurately predict the Length of Stay for each patient on case by case basis so that the Hospitals can use this information for optimal resource allocation and better functioning. The length of stay is divided into 11 different classes ranging from 0-10 days to more than 100 days.

Technical Architecture:

SOLUTION

              Hospital's management system facing various issues in bed allocation and resource allocation for heart patients because of their Length of stay(LOS) variation . Heart patients LOS may vary based on their level of severity.Inorder to overcome this issue we have to predict the patients LOS in advance. For that process we have to first collect the datasets from the e-medical records or medical records of the hospital. After collecting the datasets we have to apply any of the following algorithms Fuzzy Logic (FL), Naïve Bayes (NB),Regression Trees (CART), Random Forest (RF) and predict the length of stay of the patient and visualization for those data sets are created using EDA.