Analytics For Hospitals' Health-Care Data

Introduction

Hospital readmission is among the most critical issues in the healthcare system due to its high prevalence and cost. The improvement effort necessitates reliable prediction models which can identify high-risk patients effectively and enable healthcare practitioners to take a strategic approach. Using predictive analytics based on electronic health record (EHR) for hospital readmission is faced with multiple challenges such as high dimensionality and event sparsity of medical codes and the class imbalance. To response to these challenges, an analytical framework is proposed by data-driven approaches using hospital inpatient administrative data from a nationwide healthcare dataset.

Literature Review

[1] Data analytics for the sustainable use of resources in hospitals: Predicting the length of stay for patients with chronic diseases 2020

Identifies variables related to patients' prior admissions as important factors in the prediction of LOS in hospitals, thereby revising the current paradigm in which patients' medical histories are rarely considered for the prediction of LOS. It uses Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN).

A**dvantages:** 86 % and 91% accuracy for the COPD data set. 74 % and 85% accuracy for the pneumonia data set

Disadvantages: The prediction of LOS in other diseases needed improvement. The optimal timespan for extracting patients' historical information to obtain the best results in terms of both the predictions and time complexity of the computations needed investigation.

[2] Robust Length of Stay Prediction Model for Indoor Patients 2021.

In Selected six Machine learning (ML) models named: Multiple linear regression (MLR), Lasso regression (LR), Ridge regression (RR), Decision tree regression (DTR), Extreme gradient boosting regression (XGBR), and Random Forest regression (RFR). The selected models' predictive performance was checked using R square and Mean square error (MSE) as the performance evaluation criteria.

Advantages: Results revealed the superior predictive performance of the RFR model, both in terms of RS score (92%) and MSE score (5), among all selected models. models.

Exploratory data analysis (EDA) conclude that maximum stay was between 0 to 5 days with the meantime.

Disadvantages: Need to involve more variables in the given dataset to build a more accurate model that could predict hospital LOS more accurately.

[3] Predicting length of stay in hospitals intensive care unit using general admission features

A framework for predicting patient LOS in the ICU using different machine learning (ML) techniques are proposed.

The ML techniques used in the proposed framework are Neural Networks (NN), Classification Trees (CT), Tree Bagger (TB), Random Forest (RF), Fuzzy Logic (FL), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Regression Tree (RT) and Naive Bayes (NB).

Advantages: The best prediction accuracy was achieved by fuzzy with accuracy reach 92%, while classification tree managed to achieve a prediction accuracy of 90% coming in the second place.

Disadvantages: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[4] Using Data Analytics to Improve Hospital Quality Performance

Extracted data for 2,233,214 discharges in 2014 from 183 hospitals in the state.

Found that 20.8% of the facilities were on the quality performance frontier—20.6% of the not-for-profit facilities and 21.4% of the other facilities

Advantages: 79.2% of hospitals could improve their quality of care.

Disadvantages: As an upper bound, if all hospitals increased each quality factor performance to 100%, there would have been 11,722 (24.8%) fewer deaths, 17,840 (15.8%) fewer readmissions, and the statewide average length of stay would have been 0.71 days (13.5%) less.

[5] Big Data analytics on Diabetic Retinopathy Study (DRS) on real-time data set identifying survival time and length of stay

Using multivariate quantitative statistical method made observations on the effects of patients and hospital characteristics on diabetic in patients.

For the given set of patients, observing the study has sufficient power >=0.795 and querying to identify the co-morbidities of diabetes and the behavior of patients among the types of diabetes.

Chi-square, independent t-tests and ANOVA were used to detect the actual differences between the actual outcomes.

Advantages: The mean age of all patients was 63.72 (SD+- 13.33).

Most of the secondary diagnosis were ranged from coronary atherosclerosis (,20%) to paroxysmal ventricular tachycardia(3.4%) which includes cardiogenic shock (1.4%) to hypotension (0.3%)

Disadvantages: To put more Extensive effort into building these predictive models.

References:

[1] Data analytics for the sustainable use of resources in hospitals: Predicting the length of stay for patients with chronic diseases

https://www.sciencedirect.com/science/article/pii/S0378720619301594

[2] Robust Length of Stay Prediction Model for Indoor Patients

https://www.researchgate.net/publication/355174497_Robust_Length_of_Stay_Prediction_Model_for_Indoor_Patients

[3] Predicting length of stay in hospitals intensive care unit using general admission features

https://www.sciencedirect.com/science/article/pii/S2090447921001349

[4] Using Data Analytics to Improve Hospital Quality Performance

https://journals.lww.com/jhmonline/Fulltext/2020/08000/Using_Data_Analytics_to_Improve_Hospital_Quality.9.aspx

[5] Big Data analytics on Diabetic Retinopathy Study (DRS) on real-time data set identifying survival time and length of stay

https://www.sciencedirect.com/science/article/pii/S1877050916304926