

Risk Projection for Readmission of Congestive Heart Failure Patients on Big Data Solutions

Sreeja.D¹, N.Ananda Reddy²

¹PG Student, Dept of Computer Science And Engineering, Siddartha Educational Academy Group of Institutions,C. Gollapalli, Tirupati,AP,India

² Assistant Professor, Dept of Computer Science And Engineering, Siddartha Educational Academy Group of Institutions,C. Gollapalli,Tirupati,AP,India

e-mail: dandu.sreeja@gmail.com, cse.seat@gmail.com

*Corresponding Author: dandu.sreeja@gmail.com

Available online at: <http://www.ijcert.org>

Received: 21/10/2021,

Revised: 26/10/2021,

Accepted: 29/10/2021,

Published: 08/11/2021

Abstract:- Big Data is a collection of data that is large or complex to process using on-hand database management tools or data processing applications. It is becoming very difficult for companies to store, retrieve and process the ever-increasing data. In other words we can say, Big Data is term given to humungous amount of data which is difficult to store and process. The issue lies in using the traditional system is, how to store and analyze Big Data. Risk prediction involves integration of clinical factors with socio-demographic factors like health conditions, disease parameters, hospital care quality parameters, and a variety of variables specific to each health care provider making the task increasingly complex. Unsurprisingly, many of such factors need to be extracted independently from different sources, and integrated back to improve the quality of predictive modeling. Such sources are typically voluminous, diverse, and vary significantly over the time. This project takes Apache Hadoop, an intrinsic part for storing, retrieving, evaluating and processing huge volumes of data for processing effectively. In this work, we study big data driven solutions to predict the 30-day risk of readmission for congestive heart failure (CHF) incidents. We will predict this process by using Logistic Regression and Naive Bayes classification on the basis of data collected from patients. The results are remarkable after the comparison between the two techniques and presented through confusion matrix.

Keywords: Bigdata, Congestive heart failure, Congestive Cardiac Failure, Risk-Standardized Readmission Rate

1. Introduction

Big Data is a collection of data that is large or complex to process using on-hand database management tools or data processing applications". It is becoming very difficult for companies to store, retrieve and process the ever-increasing data. In other words we can say, Big Data is term given to humungous amount of data which is difficult to store and process. The issue lies in using the traditional system is, how to store and analyse Big Data? Apache Hadoop, which is not less than a magic bullet, has become an

intrinsic part for storing, retrieving, evaluating and processing huge volumes of data.

Heart failure (HF), often used to mean chronic heart failure (CHF), occurs the heart is unable to pump sufficiently to maintain blood flow to meet the needs of the body. The terms congestive heart failure (CHF) or congestive cardiac failure (CCF) are often used interchangeably with chronic heart failure. Symptoms commonly include shortness of breath, excessive tiredness, and leg swelling. The shortness of breath is usually worse with exercise, when lying down,

and at night while sleeping. There is often a limitation on the amount of exercise people can perform, even when well treated.

Heart failure is the leading cause of hospitalization among adults >65 years of age in the United States. Annually, >1 million patients are hospitalized with a primary diagnosis of heart failure, accounting for a total Medicare expenditure exceeding \$17 billion.

Risk prediction involves integration of clinical factors with socio- demographic factors, health conditions, disease parameters, hospital care quality parameters, and a variety of variables specific to each health care provider making the task increasingly complex. Unsurprisingly, many of such factors need to be extracted independently from different sources, and integrated back to improve the quality of predictive modelling. Such sources are typically voluminous, diverse, and vary significantly over the time.

In a recent research study, we have proposed a risk calculator tool that capable of calculating 30-day readmission risk for Congestive Heart Failure based on incomplete patient data.

In this proposed solution leverages big data infrastructure for both information extraction and predictive modelling..

2. Literature Review

A. Prediction of hospital readmission for heart failure: development of a simple risk score based on administrative data

The purpose of this study was to develop a convenient and inexpensive method for identifying an individual's risk for hospital readmission for congestive heart failure (CHF) using information derived exclusively from administrative data sources and available at the time of an index hospital discharge. Using multiple regression methods, a simple methodology was devised that segregated patients into low, intermediate and high risk for readmission. Patient characteristics, hospital features, processes of care and clinical outcomes may be used to estimate the risk of hospital readmission for CHF. However, some of the variation in rehospitalisation risk remains unexplained and may be the result of discretionary behaviour by physicians and patients.

B. Using the LACE index to predict hospital readmissions in congestive heart failure patients

The LACE index has been used to predict the risk of unplanned readmission within 30 days after hospital discharge in both medical and surgical patients. The aim of this study is to validate the accuracy of using the LACE index in CHF patients. This was a retrospective study. The LACE index score was calculated on each patient who was admitted to hospital due to an acute CHF exacerbation. Operational and clinical variables were collected from patients including basic clinical characteristics, length

of hospitalization, comorbidities, number of previous ED visits in the past 6 months before the index admission, and the number of post discharge ED revisits at 30, 60, and 90 days. All variables were analysed by multivariate logistic regression to determine the association between clinical variables and the hospital unplanned readmissions. C-statistic was used to discriminate those patients with high risk of readmissions.

C. Heart failure (HF): hospital 30-day, all-cause, unplanned risk-standardized readmission rate (RSRR) following HF hospitalization

This measure estimates a hospital-level risk-standardized all-cause unplanned 30-day readmission for patients discharged from the hospital with a principal discharge diagnosis of heart failure (HF). *This is a Centres for Medicare & Medicaid Services (CMS) only measure. The Centres for Medicare & Medicaid Services (CMS) developed the heart failure (HF) 30-day readmission measure to complement the existing HF process-of-care and mortality measures. Risk-standardized readmission rates (RSRRs) can provide important additional information about quality of care that is currently not captured by the process and mortality measures and is currently unavailable to hospitals. Variation in readmission, after adjusting for case-mix, may reflect differences in hospitals' general environments (such as coordination of care, patient safety policies, and staffing) or variation in care processes not measured in the current core measure set. Outcome measures can focus attention on a broader set of healthcare activities that affect patients' well-being. Moreover, improving outcomes is the ultimate goal of quality improvement, and thus the inclusion of outcomes measures assists in attaining improvement goals. Readmission of patients who were recently discharged after hospitalization with HF represents an important, expensive, and often preventable adverse outcome. The risk of readmission can be modified by the quality and type of care provided to these patients

D. Unplanned Readmissions after Hospital Discharge among Heart Failure Patients at Risk for 30-Day Readmission Using an Administrative Dataset and "Off the Shelf" Readmission Models"

Readmission of patients who were recently discharged after hospitalization with heart failure (HF) represents an important, expensive, and often preventable adverse outcome. The risk of readmission may be modified by the quality and type of care provided to these patients. Improving readmission rates is the joint responsibility of hospitals and clinicians. Measuring readmission will create incentives to invest in interventions to improve hospital care, better assess the readiness of patients for discharge and facilitate transitions to outpatient status. This measure is also responsive to the recent call by Medicare Payment Advisory Commission to develop readmission measures, with HF as a priority condition. Unplanned hospital readmissions within 30 days of a prior hospitalization for heart failure (HF) are

common, expensive and often preventable. These unplanned readmissions are recognized as a marker of hospital-level quality and efficiency of care and a significant contributor to rising healthcare costs. Since Heart failure is the leading cause of hospitalization among patients over the age of 65 years, the magnitude of unplanned readmissions within 30 days is enormous. Nearly one fifth of Medicare fee-for-service enrollees discharged from acute care hospitals are readmitted within 30 days, incurring additional costs of US\$17.4 billion dollars annually (1). While it is unclear whether such readmissions are entirely preventable, there is evidence that targeted interventions initiated before and/or shortly after discharge can decrease the likelihood of readmission by 25% to 45% (2-7)

3. Proposed System Design

Real world clinical data is noisy and heterogeneous in nature, severely skewed, and contains hundreds of relevant yet sometimes correlated attributes. This data resides in multiple databases such as individual EMRs, lab and imaging systems, physician notes, medical correspondences, claims, CRM systems, and hospital finance department servers. The collection, integration, and analysis of such big, complex, and noisy data in healthcare are a challenging task. For this reason, healthcare information systems can be considered as a form of big data not only for its sheer volume, but also for its complexity and diversity which makes traditional data ware housing solutions prohibitively cumbersome and ill-suited for large scale data exploration and modeling.

In this we study how a big data framework can be leveraged to extract and preprocess data. The focus of the next section will be subsequent predictive Modeling. We will leverage Hadoop as our big data framework to archive performance, scalability and fault tolerance for our task at hand. Hadoop is a popular open source map reduces implementation, which is being used as an alternative to store and process extremely large data sets on commodity hardware. Hadoop is designed to scale up from single servers to hundreds of compute nodes, each offering local computation and storage capabilities. In a recent research study, we have proposed a risk calculator tool that is capable of calculating 30-day readmission risk for Congestive Heart Failure based on incomplete patient data.

In proposed solution leverages big data infrastructure for both information extraction and predictive modeling. We study the effectiveness of our proposed solution with a Comprehensive set of experiment, considering quality and scalability. As ongoing work, we aim at leveraging big data infrastructure for our designed risk calculation tool, for designing more sophisticated predictive modeling and feature extraction techniques, and extending our proposed solutions to predict other clinical risks.

A. Architectural Diagram

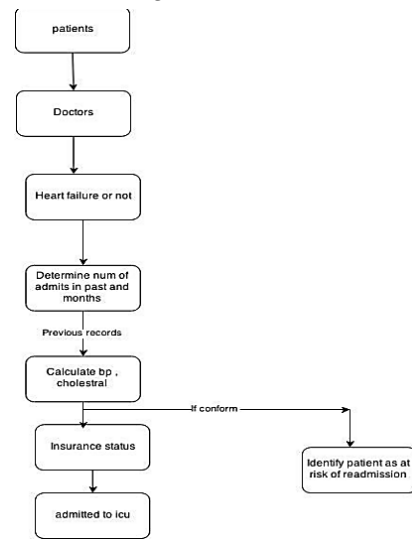


Figure 1. Flow diagram

B. Use case diagram

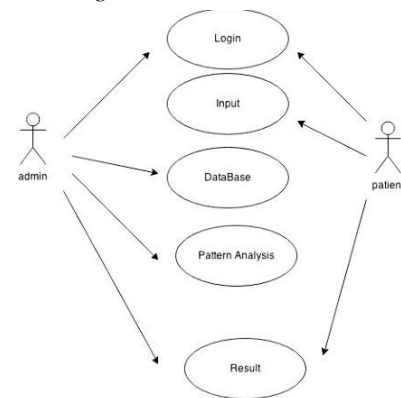


Figure 2. Use case diagram

4. Conclusion

In this work, we study the big data solution for predicting the risk of readmission for the CHF patients. Our proposed solution leverages big data infrastructure for both information extraction and predictive modeling. We study the effectiveness of our proposed solution with a comprehensive set of experiment, considering quality and scalability. As ongoing work, we aim at leveraging big data infrastructure for our designed risk calculation tool, for designing more sophisticated predictive modeling and feature extraction techniques, and extending our proposed solutions to predict other clinical risks.

References

- [1] Krumholz H. M., Normand S. L. T., Keenan P. S., Lin Z. Q., Drye E.E., Bhat K. R., Wang Y. F., Ross J. S., Schuur J. D., and Stauer B. D..Hospital 30-day heart failure readmission measure methodology.Report prepared for the Centers for Medicare & Medicaid Services.
- [2] Amarasingham R, Moore BJ, Tabak YP, Drazner MH, Clark CA, Zhang S, Reed WG, Swanson TS, Ma Y, Halm EA. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Journal of Medical Care*, 10:981-988, Feb. 2010.
- [3] An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data.
- [4] Koelling T. M., Johnson M. L., Cody R. J., and Aaronson. K. D.Discharge education improves clinical outcomes in patients with chronic heart failure. *Circulation*, 111(2):179- 185, Jan. 2005.
- [5] Impact of prior admissions on 30-day readmissions in medicare heart failure inpatients.
- [6] Meadam N., Verbiest N., Zolfaghar K., Agarwal J., Chin S., Basu Roy S., Teredesai A., Hazel D., Reed L., Amoroso P. Exploring Preprocessing Techniques for Prediction of Risk of Readmission for Congestive Heart Failure Patients. In *Data Mining and Healthcare Workshop*, in conjunction with the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2013.