Analytics Practicum

Project Proposal

Predicting 30 days readmissions of patients

for Prof. Kam Tin Seong

By Nicholas Lee Desheng, Goh Jian Hao

Table of Contents

Version History

Version	Edit description	Date
1.0	Initial draft of project proposal	27 Jan 15
2.0	Revised draft based on templates & consultation with Prof. Kam which includes the following • Detailed schedule after data cleaning consisting of univariate and bivariate analysis, data transformation	3 Feb 15
2.1	Revised draft based consultation with Prof. Kam which includes the following • Multi-collinearity analysis • Selection for variables for model construction	7 Feb 15

Project Motivation

Hospitals have been studying about the likelihood of patients readmitting within 30 days starting on the day of discharge primarily to reduce costs and operational overhead. It has grown to be a governmental concern as health systems, especially in UK, has decided to incentivize hospitals that abide by the rules and penalize those who did not manage to reduce their number of 30 days readmissions.

Furthermore, for governments with welfare systems that provide for health care, patients readmitted within 30 days may be an avoidable expense if the hospitals were able to identify such patients during the first diagnosis.

There are already several studies that analyze the rate and causes of readmissions, with the period of analysis ranging from 14 to 60 days. These studies looked for traits and indicators that may predict a patient's likelihood of readmission. The predictor variables used varied greatly amongst the different studies. Some used the patient's demographic indicators, such as weight, age, race, while others used the diagnosis for admission and length of stay. Given the binary nature of the outcomes (a patient is either readmitted or not), many of these models used logistic regression to identify predictor variables. However, no definitive model has emerged due to the aforementioned variables that differed between each study.

At the same time, various researchers claimed their superiority over other studies. Even though LACE (Length of stay, Acuity of admission, Comorbidity, Emergency department visits) index has been acknowledged as the gold standard in prediction of 30 day readmissions rates, authors claimed that their model perform better, often with caveats.

The outcomes of such a study allow hospitals to identify patients with a higher risk of readmissions and prescribe interventions in the form of house visits, or a wide-spectrum treatment in order to mitigate the problem. Data analytics, especially predictive ones, enable hospitals to do so. If effectively implemented, hospitals can reduce their costs and focus their limited resources to prevent avoidable readmissions. Consequently, hospitals may be able to stretch their resources to care for a larger number of patients, instead of serving readmitted patients..

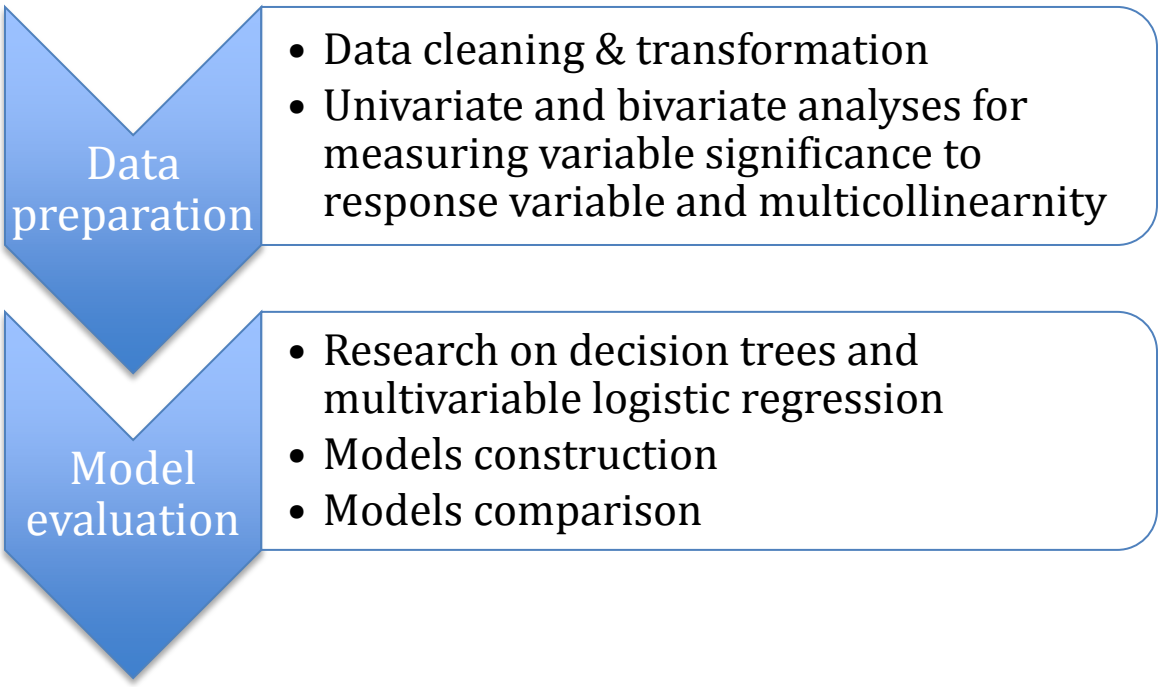
Objective

To derive a model predicting the likelihood of a patient readmitting within 30 days using 2 approaches (i.e. random forest decision tree & multivariable logistic regression), along with its corresponding ROC curves to determine the predictive power of our models, thereby culminates in an evaluation of models.

Stakeholders

Supervisor & Sponsor: Prof. Kam Tin Seong (Associate Professor of Information Systems; Senior Advisor, SIS Programmes in Analytics)

Scope of Work



Technical Complexities

- Understanding and applying principles of methods (e.g. multivariable logistic regression, decision trees) to derive models
- Understanding and applying statistical mathematics for multivariable analysis (e.g. receiver operating characteristic as a predictive performance metric of models)
- Employing gainful usage of SAS software (Enterprise Miner, JMP) and open source programs like Tableau Public

Schedule

Week	Activities	Milestone
3	• Review literature (10 selected papers) • Proposal	Proposal

	● Sense-making through data pre-processing	
4	● Further data pre-processing ● Preliminary understanding through exploration in SAS software	
5	● Research/literature review on the mechanics & mathematics of models (i.e. random forest DT & multivariable logistic regression) ● Univariate & bivariate analysis: attempting to distinguish highly-correlated variables leading to readmission within 30 days using SAS EM	
6	● Preliminary construction of DT structure ● Univariate & bivariate analysis: attempting to distinguish highly-correlated variables leading to readmission within 30 days using SAS EM (continued) ● Selection of variables for model construction	
7	● Slides deck drafting + rehearsal ● Report drafting	Midterm presentation + report
8	● Model construction: Further work on the 2 models (i.e. decision trees & logistic regression) ○ Research/literature review of various sub-type of decision trees and applicability ○ Understanding the inner workings of logistic regressions	
9	● Model testing against test data to obtain predictive power values (i.e. Area Under Curve in receiver operating characteristic curves)	
10	● Explore agent-based modelling for better understanding	
11	● Results compilation	
12	● Slides deck drafting + rehearsal ● Reporting drafting	
13	● Poster drafting	
14	● Final preparations	Final presentation + report + poster

References

Billings, J., Dixon, J., Mijanovich, T. and Wennberg, D. (2006) "Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients," British Medical Journal 333(7563), 327-333.

Vaithianathan, R., Jiang, N. and Ashton, T. (2012) "A Model for Predicting Readmission Risk in New Zealand." Auckland University of Technology, Working Paper Series

Lemke, K. (2012) "Building a Predictive Model for 30-Day Inpatient Readmission using PROC PHREG" NESUG

Yo, I. et al (2012) "Factors Affecting Unplanned Readmissions from Community Hospitals to Acute Hospitals: A prospective Observational Study" Ann Acad Med Singapore 2009;38:113-20

Donze, J., Aujesky, D., Williams, D., Schnipper, J. (2013) "Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients" JAMA Intern Med. 2013:173(8)632-628

Billings, J., Blunt, I., Steventon, A. et al. (2012) "Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30)" BMJ Open 2012;00:e001667. doi:10.1136/bmjopen-2012-001667

Yu, S.P., van Esbroeck, A., Farooq, F. et al. (2013) "Predicting readmission risk with institution specific prediction models" In Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI'13). IEEE, 551–556.

Strack, B., DeShazo, J, P., Gennings, C. et al. (2014) "Impact of HbA1c measurement on hospital readmission rates: analysis of 70, 000 clinical database patient records" Biomed Research International, Vol 2014, ID: 781670, 11 pages

Natale, J., Wang, S, Y., Taylor, J. (2013) "A decision tree model for predicting heart failure patient readmissions" In proceedings of the 2013 Industrial and Systems Engineering Research Conference

Tan, S, Y., Low, L, L., Yong, Y. et al. (2013) "Applicability of a previously validated readmission predictive index in medical patients in Singapore: a retrospective study" BMC Health Services Research, 13:366 at www.biomedcentral.com/1472-6963/13/366