Heinz 95-845: Project Proposal

Yue Wu   yuew5@andrew.cmu.edu
Heinz College
Carnegie Mellon University
Pittsburgh, PA, United States

Cheng Chen   chengc3@andrew.cmu.edu
Engineering & Technology Innovation Management
Carnegie Mellon University
Pittsburgh, PA, United States

1. Proposal Details

1.1 What is your proposed analysis? What are the likely outcomes?

The medications for hospitalized diabetes patients vary from patient to patient and time to time. During hospitalization, patients are usually diagnosed several times and their prescriptions often change accordingly. We plan to explore the impact of insulin medication on early hospital readmission rates by analyzing the dataset extracted from Health Facts database which represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. We define the outcome as having two possible values: one is for the patient was readmitted within 30 days of discharge, another is for no readmission within 30 days[1].

1.2 Why is your proposed analysis important?

Diabetes usually comes with a lot of other complications, including hypertension, hyperlipidemia, cardiovascular disease, kidney disease and even cancers. For the diabetes patients, readmission is one of the few things that they want to avoid during the treatment process. By analyzing the factors which may affect the rate of readmission, it will increase the effectiveness and safety to the diabetes patients as well reduce the amount of further work of physicians.

1.3 How will your analysis contribute to existing work? Provide references, *e.g.*, see: Name and Name (2016).

Previous works have proven some important practices can reduce 30-day readmission rates for diabetes patients. For instance, Beata Strack and Jonathan P. DeShazo et al.(2014) address the decision to test for HbA1c, rather than the values of the HbA1c result could affect the readmission possibility in the study of Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient[2]. Daniel J.Rubin et al.(2014) suggest that several interventions may reduce the risk of early readmission for patients with diabetes, including inpatient diabetes education, improving communication of discharge instructions, and involving patients more in medication reconciliation and post-discharge planning in the study of Early readmission among patients with diabetes:

A qualitative assessment of contributing factors[3]. Our study will study the impact of the change in dosage of insulin on the probability of readmission, which could extend the research on exploring best practices to reduce the readmission rates for diabetes patients.

1.4 Describe the data. If applicable, please also define Y outcome(s), U treatment, V covariates, and W population.

The dataset is extracted from Health Facts database which represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes and almost 70,000 inpatient diabetes encounters. It contains valuable but heterogeneous and difficult data in terms of missing values, incomplete or inconsistent records, and high dimensionality understood not only by number of features but also their complexity[4]. In our study, the Y outcome is the early hospital readmission for diabetes patients, the U treatment is the insulin medication, the V covariates include demographics, results of diagnosis and test, type of admission, other medications and so on, and the W population is diabetic inpatients with laboratory tests and medications during the encounter.

1.5 What evaluation measures are appropriate for the analysis? Which measures will you use?

We can make use of ROC curve, confusion matrix, F1 score to evaluate the model we built. Furthermore, we will use level of significance to evaluate the impact of treatment.

1.6 What study design, pre-processing, and machine learning methods do you intend to use? Justify that the analysis is of appropriate size for a course project.

We will conduct ETL and preliminary analysis on the raw dataset, including dealing with missing data. For the feature engineering, we may use Principal Component Analysis to reduce dimensionality and try to control randomization as far as possible. Then, we will use logistic regression as our primary machine learning methods and design low-bias tests for significance of variables. Logistic regression will be used to measure the relationship between the change of insulin dosage and early readmission while controlling for other covariates such as demographics, type of admission, other medications and so on. This study will be limited by a nonrandomized study design.

1.7 What are possible limitations of the study?

As discussed earlier, large clinical databases usually contains many missing values, incomplete or inconsistent records, which may introduce uncertainty in our study. Also, it is limited by a non-randomized study design. As we cannot obtain a non-randomized dataset, we cannot implement RCTs in this study.

1.8 Who will your analytic pipeline? In one or two sentences, describe an example of its use.

All medical workers such as physicians, doctors and pharmacist can use this analysis as a reference to adjust their treatment for diabetes patients in the future.

References

Her Name and His Name. Aamlp project. In *The Journal Where It Was Published*, 2016.

[1] Nathan, David M., and DCCT/Edic Research Group. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. In *diabetes care 37.1: 9-16*, 2014.

[2] Beata Strack, Jonathan P. DeShazo and Chris Gennings et al. Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records" In *BioMed Research International*, 2014.

[3] Daniel J.Rubin and KellyDonnell-Jackson et al. Early readmission among patients with diabetes: A qualitative assessment of contributing factors. In *Journal of Diabetes and its Complications*, 2014.

[4] K. J. Cios and G. W. Moore. Uniqueness of medical data mining. In *Artificial Intelligence in Medicine*, 2002.