

## Title

Predicting early readmissions for diabetes patients using clinical records from 130 US hospitals

## Problem

In 2012, patients with diabetes incurred approximately \$124 billion in annual expenditure for hospital care in the United States. Although diabetic patients represent about 8% of the US population, they account for 23% of hospitalizations (8.8 million) each year. Among all the healthcare utilizations, hospitalization is major cost driver for diabetes patients. Readmission within 30 days after a hospital discharge is a high-priority healthcare quality measure and target for cost reduction. It is a more severe situation among patients admitted for diabetes as the rate in those patients is 14.4–21.0%<sup>1</sup>, as compared to the overall early readmission rate, which is 8.5–13.5%<sup>2,3</sup>. Therefore, this study aims to build a machine learning model to precisely predict the likelihood of early readmissions among inpatient diabetes patients using a large hospital database from 1999 to 2008.

## Who might care?

Physicians, healthcare researchers, and health plan insurance companies might be interested in using this model to predict the readmission outcome for diabetic patients. Based on the predicted probability of early readmission, early interventions can be done to patients such as discharge education, further lab testing to ensure stabilized glucose level, and home care reminders of medications.

## Data

The [hospital dataset](#) for this study was originally obtained from the Health Facts database, which was an extract representing 10 years (1999–2008) of clinical care at 130 hospitals and integrated delivery networks throughout the United States. It was later adjusted to suit for a diabetes research and was released on behalf of the Center for Clinical and Translational Research, Virginia Commonwealth University on UCI Repository<sup>4</sup>. The adjusted data contains 100000 of encounters with 55 features representing patient and hospital characteristics. All encounters in the dataset are for diabetes admissions.

**Modeling approach:** Since we want to predict the likelihood of an early readmission, a supervised classification algorithm is a perfect choice to build the predictive model. The

---

<sup>1</sup> Rubin, D. J., E. Handorf, and M. McDonnell. "Predicting early readmission risk among hospitalized patients with diabetes (7796)." *ENDO* (2013): 95th.

<sup>2</sup> "Hospital Readmissions in Pennsylvania 2010". Pennsylvania Health Care Cost Containment Council. Accessed 21/06/2018

<sup>3</sup> Friedman, Bernard, H. Joanna Jiang, and Anne Elixhauser. "Costly hospital readmissions and complex chronic illness." *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* 45.4 (2008): 408-421.

<sup>4</sup> Beata Strack, Jonathan P. DeShazo, Chris Gennings, Juan L. Olmo, Sebastian Ventura, Krzysztof J. Cios, and John N. Clore, "Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records," *BioMed Research International*, vol. 2014, Article ID 781670, 11 pages, 2014.

classification algorithm not only classifies classes (in our case, two classes: “early-readmitted” and “not early-readmitted”) but also predict the probability of each class. This data set is supposed to have imbalanced classes (only 10-15% of the patients will be re-admitted), which makes the project challenging. We plan to use different approaches to tackle the imbalanced class issue and choose the best one. Also, we will try various classification algorithms and pick the one which performs best.