Comprehensive Analysis of Diabetic Readmissions

Milestone: Final Project Proposal

Group 18

Mallika Gaikwad
Divya Babulal Shah

gaikwad.mal@northeastern.edu
shah.divyab@northeastern.edu

Percentage of Effort Contributed by Student 1: 50%

Percentage of Effort Contributed by Student 2: 50%

Signature of Student 1: Mallika Gaikwad

Signature of Student 2: Divya Babulal Shah

Submission Date: 2nd February 2024

Problem Setting :

Diabetes, a chronic and prevalent health issue, poses significant challenges in healthcare, leading to higher rates of hospital readmissions. Hospital readmission, the return of a patient within a specific timeframe after discharge, is particularly common among those with diabetes. The frequency of readmissions serves as a critical metric for evaluating hospital quality and the effectiveness of diabetes management.

Problem Definition:

The project aims to leverage data mining techniques to analyze diverse medical data related to diabetic patients. Specifically, the goal is to predict whether a diabetic patient will be readmitted to the hospital based on various factors, including age, gender, admission type, medications administered, duration of hospitalization, laboratory test results.

Data Sources:

The primary data source for this project is UCI Machine Learning Repository, from which this dataset is obtained (Dataset Link).

Data Description:

The dataset spans a significant temporal window from 1999 to 2008, originating from a comprehensive network of 130 US hospitals. Comprising over 100,000 patient records, this dataset encompasses a variety of information, featuring a total of 50 attributes. Within this dataset, 13 are numerical, and 47 are categorical, offering a multi-dimensional perspective on clinical care. Noteworthy among the features are 23 indicators detailing various medications shedding light on prescription specifics and dosage changes. The dataset is comprehensive, covering a spectrum of attributes like patient identification numbers, race, gender, age, admission types, physician specialties, lab test metrics, HbA1c test results, diagnoses, medication counts, diabetic prescriptions, and counts of outpatient, inpatient, and emergency visits in the year preceding

hospitalization. The target column to the analysis is the 'readmitted' column, offering insights into patient readmission patterns. With values denoting readmission within "<30" days, ">30" days, or "NO" record, this dataset provides a robust foundation for exploring and understanding the dynamics of diabetic patient readmissions over the specified period.

The features/variables identified as crucial in predicting our response variable – 'readmitted' are as follows:

1. **Age:**

   Age is a key demographic variable that often correlates with health outcomes. It helps in understanding age-specific trends in readmission rates, as different age groups may experience unique healthcare challenges.

2. **Diag_1 and Diag_2 (Primary and Secondary Diagnoses):**

   These features provide insights into the primary and secondary diagnoses based on the first three digits of the ICD9 coding system. Understanding specific medical conditions is essential for predicting readmission risks and tailoring patient care plans accordingly.

3. **Admission_source_id:**

   The 'admission_source_id' indicates where the patient was referred from or how they arrived at the hospital. This information is vital for assessing the context of hospitalization, such as planned admissions through physician referrals or urgent admissions via the emergency room.

4. **Race:**

   The 'race' variable is vital for uncovering healthcare disparities, influencing access to care, treatment outcomes, and readmission rates. Analyzing it provides insights into potential inequalities, guiding targeted interventions in diverse patient populations.

5. **Discharge_disposition_id:**

   The 'discharge_disposition_id' is vital, revealing outcomes post-hospital discharge with codes for scenarios like returning home or transfer. It is crucial for understanding recovery patterns, aiding in predicting diabetic patient readmission risks.

In summary, in addition to the above-mentioned variables, various numerical indicators may also play important roles. Further statistical analysis and modeling are needed to determine the precise significance of each variable in predicting readmissions.