Predicting hospital readmission using a clinical relational database

Project Step 2: Outline

Data

ABOUT THE DATASET

1. MIMIC-III database comprises de-identified health-related data associated with patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012.
2. It contains 53,423 distinct hospital admissions for adult patients (aged 16 years or above) admitted to critical care units between 2001 and 2012. In addition, it contains data for 7870 neonates admitted between 2001 and 2008.
3. Includes information such as demographics, vital sign measurements made at the bedside (~1 data point per hour), laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality (including post-hospital discharge).[1]
4. MIMIC-III is a relational database consisting of 26 tables with roughly the following categories:
    - 4.1. Charted events: notes, laboratory tests, and fluid balance are stored in a series of 'events' tables. OUTPUTEVENTS table, LABEVENTS table
    - 4.2. Dictionary tables: provide definitions for identifiers. CHARTEVENTS table
    - 4.3. Patient tracking tables: define and track patient stays: ADMISSIONS; PATIENTS; ICUSTAYS; SERVICES; and TRANSFERS
    - 4.4. Cross-referencing dictionary tables: D_CPT; D_ICD_DIAGNOSES; D_ICD_PROCEDURES; D_ITEMS; and D_LABITEMS.
    - 4.5. Other tables: contain data associated with patient care, such as physiological measurements, caregiver observations, and billing information.

RELATED INFORMATION - STATISTICS, ACCESS TO THE DATASET, AND WORKFLOW

1. This is an accurately depicted relational database. [2]
2. The median age of adult patients is 65.8 years (Q1–Q3: 52.8–77.8), 55.9% of patients are male, and in-hospital mortality is 11.5%. The median length of an ICU stay is 2.1 days (Q1–Q3: 1.2–4.6) and the median length of a hospital stay is 6.9 days (Q1-Q3: 4.1–11.9). [3]
3. Accessing this dataset requires completing a CITI training course and requesting on the PhysioNet account. Dataset link: [1]
4. For training purposes, tables (mainly ADMISSIONS and NOTEEVENTS) of the relational database are merged. According to what we had discussed in the meeting, there is a need of developing a data reader model that can merge tables and provide labels as a concrete dataset. As of now, out of the huge chunk of the database, initially, we have considered 60 data rows of patients and their information.

References:

1. https://physionet.org/content/mimiciii/1.4/
2. https://miro.medium.com/max/1838/1*H_bDafx8LJ98CsV2hQP1OQ.png
3. https://www.nature.com/articles/sdata201635