

PRJ-002: Diabetes 130-US Hospitals for Years 1999-2008

The dataset represents ten years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. Each row concerns hospital records of patients diagnosed with diabetes, who underwent laboratory, medications, and stayed up to 14 days. The goal is to determine the early readmission of the patient within 30 days of discharge. The problem is important for the following reasons. Despite high-quality evidence showing improved clinical outcomes for diabetic patients who receive various preventive and therapeutic interventions, many patients do not receive them. This can be partially attributed to arbitrary diabetes management in hospital environments, which fail to attend to glycemic control. Failure to provide proper diabetes care not only increases the managing costs for the hospitals (as the patients are readmitted) but also impacts the morbidity and mortality of the patients, who may face complications associated with diabetes.

Dataset Characteristics Multivariate

Subject Area Health and Medicine

Associated Tasks Classification, Clustering

Feature Type Categorical, Integer

Instances 101766

Features 47

Dataset Information

What do the instances in this dataset represent?

The instances represent hospitalized patient records diagnosed with diabetes.

Are there recommended data splits?

No recommendation. The standard train-test split could be used. Can use three-way holdout split (i.e., train-validation-test) when doing model selection.

Does the dataset contain data that might be considered sensitive in any way?

Yes. The dataset contains information about the age, gender, and race of the patients.

Additional Information

The dataset represents ten years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Information was extracted from the database for encounters that satisfied the following criteria.

- (1) It is an inpatient encounter (a hospital admission).
- (2) It is a diabetic encounter, that is, one during which any kind of diabetes was entered into the system as a diagnosis.

- (3) The length of stay was at least 1 day and at most 14 days.
- (4) Laboratory tests were performed during the encounter.
- (5) Medications were administered during the encounter. The data contains such attributes as patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab tests performed, HbA1c test result, diagnosis, number of medications, diabetic medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc.