Prediction of hospital readmissions for patients with diabetes

Neeraj Kumar Reddy Panta

Ruchi Dilip Kukde

Data Set

- **Multivariate data to analyze factors related to hospital readmission as well as other outcomes pertaining to patients with diabetes**

- **Origin:**

The data was collected in the form of comprehensive clinical records across hospitals throughout United States by Health Facts database – Cerner Corporation, Kansas City

- **Source:**

The data was submitted to UCI Machine Learning Repository [1] in 2014 on behalf of the Center for Clinical and Translational Research, Virginia Commonwealth University, a recipient of NIH CTSA grant UL1 TR00058 and a recipient of the CERNER data [2]

Data Set

- **File Name:** diabetic_data.csv and IDs_mapping.csv
- **File format:** Comma separated values (csv)

- **Data Set Information:**
- The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks
- It includes 55 features representing patient and hospital outcomes

Data Set

Information was extracted from the database for encounters that satisfied the following criteria:

- It is an inpatient encounter (a hospital admission).
- It is a diabetic encounter, that is, one during which any kind of diabetes was entered to the system as a diagnosis.
- The length of stay was at least 1 day and at most 14 days.
- Laboratory tests were performed during the encounter.
- Medications were administered during the encounter.

Data Dictionary

S.NO	Field Name	Field Type	Description and Values
1	Encounter ID	Numeric	Unique identifier of an encounter
2	Patient number	Numeric	Unique Identifier of a patient
3	Race	Character	Values: Caucasian, Asian, African American, Hispanic, and other
4	Gender	Character	Values: male, female, and unknown/invalid
5	Age	Character	Grouped in 10-year intervals: [0, 10), [10, 20) ..., [90, 100)
6	Weight	Numeric	Weight in pounds
7	Admission Type	Character	Integer identifier corresponding to 8 distinct values, for example, emergency, urgent, elective, newborn, and not available
8	Discharge Disposition	Character	Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, discharged/transferred to another type of inpatient care institution and not available
9	Admission source	Character	Integer identifier corresponding to 26 distinct values, for example, physician referral, clinic referral, court/law enforcement, emergency room, and transfer from a hospital
10	Time in Hospital	Numeric	Integer number of days between admission and discharge
11	Payer code	Character	Integer identifier corresponding to 23 distinct values, for example, Blue Cross\BlueShield, Medicare, and self-pay
12	Medical Specialty	Character	Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family\general practice, and surgeon
13	Number of lab procedures	Numeric	Number of lab tests performed during the encounter
14	Number of procedures	Numeric	Number of procedures (other than lab tests) performed during the encounter
15	Number of medications	Numeric	Number of distinct generic names administered during the encounter
16	Number of outpatient visits	Numeric	Number of outpatient visits of the patient in the year preceding the encounter
17	Number of emergency visits	Numeric	Number of emergency visits of the patient in the year preceding the encounter
18	Number of inpatient visits	Numeric	Number of inpatient visits of the patient in the year preceding the encounter

Data Dictionary

S.NO	Field Name	Field Type	Description and Values
19	Diagnosis 1	Character	The primary diagnosis (coded as first three digits of ICD9); 848 distinct values
20	Diagnosis 2	Character	Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values
21	Diagnosis 3	Character	Additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct values
22	Number of Diagnoses	Numeric	Number of diagnoses entered to the system
23	Glucose serum test result	Character	Indicates the range of the result or if the test was not taken. Values: “>200,” “>300,” “normal,” and “none” if not measured
24	Alc test result	Character	Indicates the range of the result or if the test was not taken. Values: “>8” if the result was greater than 8%, “>7” if the result was greater than 7% but less than 8%, “normal “if the result was less than 7%, and “none” if not measured.
25	Change of medications	Character	Indicates if there was a change in diabetic medications (either dosage or generic name). Values: “change” and “no change”
26	Diabetes medications	Character	Indicates if there was any diabetic medication prescribed. Values: “yes” and “no”
27	24 features for medications	Character	For the generic names: metformin, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, sitagliptin, insulin, glyburide-metformin, glipizide-metformin, glimepiride-pioglitazone, metformin-rosiglitazone, and metformin-pioglitazone, the feature indicates whether the drug was prescribed or there was a change in the dosage. Values: “up” if the dosage was increased during the encounter, “down” if the dosage was decreased, “steady” if the dosage did not change, and “no” if the drug was not prescribed
28	Readmitted	Character	Days to inpatient readmission. Values: “<30” if the patient was readmitted in less than30 days, “>30” if the patient was readmitted in more than 30 days, and “No” for no record of readmission.

References

[1]https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008

[2] 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.