Case-Study 21: Diabetes 130-US hospitals for years 1999-2008 Dataset

**Overview**

This dataset consists of risk factors related to hospital readmission of diabetes patients collected from 130 hospitals and integrated delivery networks from 1999-2008. Examples of attributes including race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab test performed, HbA1c test result, diagnosis, number of medication, diabetes medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization.

**Driving Challenges**:

1. Can we develop risk predictive models using available risk factors associated with 30-day readmission rate of diabetes patients?
2. Is polypharmacy (number of medications received) importance as a risk factor for hospital readmission?
3. What are the combination of health conditions (comorbidity) that greater impact the hospital readmission rate?
4. Is there a relationship between comorbidity and diabetes severity as measured by HbA1C level?

**Meta-data**: Number of subjects: 100,000 Time: 1999-2008

List of features and their descriptions

* Encounter ID: Unique identifier of an encounter
* Patient number: Unique identifier of a patient
* Race: AfricanAmerican, Asian, Caucasian, Hispanic, Other, ? (Missing data)
* gender: Female, Male, Unknown/Invalid

- age: 30-40, 40-50, 50-60, 60-70, 70-80, other

- weight: 25-50, 50-75, 75-100, 100-125, 125-150, other

* admission\_type: integer identifier corresponding to 9 distinct values, for examine emergency, urgent elective, newborn, and not available
* discharge\_disposition: integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available
* admission\_source: integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital
* time\_in\_hospital: integer number of days between admission and discharge
* payer\_code: integer identifier corresponding to 23 distinct values, for example, Blue Cross/Blue Shield, Medicare, and self-pay
* medical specialty: integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, InternalMedicine, Emergency/Trauma, Family/GeneralPractice, Cardiology, Surgery-General, Other
* numb\_lab\_procedures: number of lab test performed during the encounter
* num\_procedures: number of procedures (other than lab tests) performed during the encounter
* num\_medications: number of distinct generic names administered during the encounter
* number\_outpatient: number of outpatient visits of the patient in the year preceding the encounter
* number\_emergency: number of emergency visits of the patient in the year preceding the encounter
* number\_inpatient: number of inpatient visits of the patient in the year preceding the encounter
* diag\_1: the primary diagnosis (coded as first three digits of ICD9); 848 distinct values
* diag\_2: secondary diagnosis (coded as first three digits of ICD9); 923 distinct values
* diag\_3: additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct values
* number\_diagnoses: number of diagnoses entered to the system
* max\_glu\_serum: glucose serum test result, indicates the range of the result or if the test was not taken (“>200”, “>300”, “Normal”, and “None” if not measured)
* A1Cresult: indicate 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 results was less than 7%, and “none” if not measured.
* Change: Change of medication, indicates if there was change in diabetic medications (either dosage or generic name). Values: “change” and “no change”
* diabetesMed: diabetes medications: indicates if there was any diabetic medication prescribed. Values: “yes” and “no”.
* 24 features for diabetes medications for the following generic names. The feature indicates whether the drug was prescribed or there was a change in 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
  + Metformin: Down, No, Steady, Up
  + Repaglinide Down, No, Steady, Up
  + Nateglinide Down, No, Steady, Up
  + Chlorpropamide Down, No, Steady, Up
  + Glimepiride Down, No, Steady, Up
  + Acetohexamide Down, No, Steady, Up
  + Glipizide Down, No, Steady, Up
  + Glyburide Down, No, Steady, Up
  + Tolbutamide Down, No, Steady, Up
  + Pioglitazone Down, No, Steady, Up
  + Rosiglitazone Down, No, Steady, Up
  + Acarbose Down, No, Steady, Up
  + Miglitol Down, No, Steady, Up
  + Troglitazone Down, No, Steady, Up
  + Tolazamide Down, No, Steady, Up
  + Examine Down, No, Steady, Up
  + Citoglipton Down, No, Steady, Up
  + Insulin Down, No, Steady, Up
  + Glyburide.metformin Down, No, Steady, Up
  + Glipizide.metformin Down, No, Steady, Up
  + Glimepiride.pioglitazone Down, No, Steady, Up
  + Metformin.rosiglitazone Down, No, Steady, Up
  + Metformin.pioglitazone Down, No, Steady, Up
  + Readmitted: days to inpatient readmission. Value “<30” if the patient was readmitted in less than 30 days, “>30” if the patient was readmitted in more than 30 days, and “NO” for no record of readmission.

**Data Provenance**:

This dataset is retrieved from the Health Facts database (Cerner Corporation, Kansas City, MO), a national data warehouse that collects comprehensive clinical records across hospitals throughout the United States. This de-identified data were submitted on behalf of the Center for Clinical and Translational Research, Virginia Commonwealth University to the Machine Learning Repository of University of California Irvine.

Link to complete dataset

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

**Citation**:

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.