

Dataset Name: Diabetes 130-US hospitals for years 1999-2008 Data Set

Dataset Source: <http://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008#>

Topic: Healthcare

Project Idea No.1:

Goal: (Regression) Predict Time in hospital variable which is an integer number of days between admission and discharge.

Business view: If we can predict time in hospital, we can optimize hospital rooms usage, so we will have available rooms for all incoming patients.

Project Idea No.2:

Goal: (Classification) Predict Change of medications variable which Indicates if there was a change in diabetic medications (either dosage or generic name). Values: “**change**” and “**no change**”.

Business view: Hospital can create an online predictor for patients, and after inserting their data, if they are classified into “**Change**” medication, they need to go to hospital to get checked for further diagnosis and prescription.

Variable summary: Check the appendix.

Project Plans:

1. Data exploration
2. Data cleaning
3. Independent variable selection
4. Descriptive analysis
5. Predictive model selection
6. Result and recommendation

Appendix

Variable Summary

Feature name	Type	IV/DV	Description and values	% missing
Encounter ID	Primary Key	IV	Unique identifier of an encounter	0%
Patient number	Primary Key	IV	Unique identifier of a patient	0%
Race	Nominal	IV	Values: Caucasian, Asian, African American, Hispanic, and other	2%
Gender	Nominal	IV	Values: male, female, and unknown/invalid	0%
Age	Interval	IV	Grouped in 10-year intervals: 0, 10), 10, 20), ..., 90, 100)	0%
Weight	Ratio	IV	Weight in pounds.	97%
Admission type	Nominal	IV	Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, newborn, and not available	0%
Discharge disposition	Nominal	IV	Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available	0%
Admission source	Nominal	IV	Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital	0%
Time in hospital	Ratio	DV	Integer number of days between admission and discharge	0%
Payer code	Nominal	IV	Integer identifier corresponding to 23 distinct values, for example, Blue Cross/Blue Shield, Medicare, and self-pay	52%
Medical specialty	Nominal	IV	Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family/general practice, and surgeon	53%
Number of lab procedures	Ratio	IV	Number of lab tests performed during the encounter	0%
Number of procedures	Ratio	IV	Number of procedures (other than lab tests) performed during the encounter	0%
Number of medications	Ratio	IV	Number of distinct generic names administered during the encounter	0%
Number of outpatient visits	Ratio	IV	Number of outpatient visits of the patient in the year preceding the encounter	0%

Feature name	Type	IV/DV	Description and values	% missing
Number of emergency visits	Ratio	IV	Number of emergency visits of the patient in the year preceding the encounter	0%
Number of inpatient visits	Ratio	IV	Number of inpatient visits of the patient in the year preceding the encounter	0%
Diagnosis 1	Nominal	IV	The primary diagnosis (coded as first three digits of ICD9); 848 distinct values	0%
Diagnosis 2	Nominal	IV	Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values	0%
Diagnosis 3	Nominal	IV	Additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct values	1%
Number of diagnoses	Ratio	IV	Number of diagnoses entered to the system	0%
Glucose serum test result	Nominal	IV	Indicates the range of the result or if the test was not taken. Values: ">200," ">300," "normal," and "none" if not measured	0%
A1c test result	Nominal	IV	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.	0%
Change of medications	Nominal	DV	Indicates if there was a change in diabetic medications (either dosage or generic name). Values: "change" and "no change"	0%
Diabetes medications	Nominal	IV	Indicates if there was any diabetic medication prescribed. Values: "yes" and "no"	0%
24 features for medications	Nominal	IV	For the generic names. 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	0%
Readmitted	Nominal	IV	Days to inpatient readmission. Values: "<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.	0%