



Big Data Analytics for Competitive Advantage

Deliverable - 1

Diabetic Patient Readmission Prediction

ITCS-6100

Submitted by:

Ajith Kumar Gannamaneni

801305442

agannama@uncc.edu

Vivek Sai Gujja

801305448

vgujja@uncc.edu

Abhinav Botla

801290454

abotla@uncc.edu

Akhil Rayapati

801298428

arayapat@uncc.edu

Communication plan to include project artifact repository:

A word file will be uploaded in canvas in order to accommodate Deliverable-1. Once the other deliverables are met, a repository will be created containing code and screenshots of estimated accuracy, Recall value and column impact. Public access will be given for the repository.

Github Repository - <https://github.com/raojith2210/Diabetic-Patient-Readmission-Prediction.git>

Business Problem:

Hospital readmission is a significant factor in overall medical costs and is a developing metric for care quality. Diabetes, similar to other chronic medical conditions, is associated with increased risk of hospital readmission. Hospitals are striving harder to get the patient details in order to equip themselves better and serve the patients effectively. This can be considered as a business problem in this case. In this model, we will predict whether the high-risk diabetic-patients is likely to get readmitted to hospital after previous encounter within thirty days or after thirty days or totally not.

Domain Knowledge:

Original data source is stored at [UCI Repository](#). This data has been prepared to analyse factors related to readmission as well as other outcomes pertaining to patients with diabetes.

Input Data Set:

We have considered a dataset describing the patient details from [Amazon Sage maker](#).

The dataset represents clinical treatment provided over a ten-year period (1999–2008) at 130 US hospitals and integrated delivery networks. More than 50 features that represent patient and hospital outcomes are available. For interactions that met the following requirements, data was taken out of the database.

- This is a hospital visit (a hospital admission).
- It is a diabetic encounter, that is, one in which a diagnosis of diabetes of any kind was made.
- The visit was at least one day long and no longer than fourteen days.
- Laboratory tests were performed during the encounter.
- 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 test performed, HbA1c test result, diagnosis, number of medications, diabetic medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc. Below tabular format depicts the metadata of each attribute in the table:

Column Name	Data type	Data type
race	STRING	Caucasian, Asian, African American or Hispanic
time_in_hospital	INT	Number of days between admission and discharge a.k.a. length of stay
number_outpatient	INT	Number of outpatient visits of the patient in a given year before the encounter
number_inpatient	INT	Number of inpatient visits of the patient in a given year before the encounter
number_emergency	INT	Number of emergency visits of the patient in a given year before the encounter
number_diagnoses	INT	Number of diagnoses entered in to the system
num_procedures	INT	Number of procedures (other than lab tests) performed during the encounter
num_medications	INT	Number of distinct generic medicines administrated during the encounter
num_lab_procedures	INT	Number of lab tests performed during the encounter
max_glu_serum	STRING	Indicates the range of result or if the test was not taken. Values: ">200", ">300", "normal" and "none" - if not measured
gender	STRING	Values: "Male", "Female" and "Unknown/Invalid"
diabetes_med	INT	Indicates if any diabetes medication was prescribed.
change	STRING	Indicates if there was a change in diabetic medications (ether dosage or generic name). Values: "change" or "no change"
age	INT	Age of patient at the time of encounter
a1c_result	STRING	Indicates range of the result of if it was not taken. Values: ">8", ">7", "normal" and "none"
readmitted	STRING	Days to inpatient readmission. Values: "<30" if patient readmitted less than 30 days, ">30" if patient readmitted after 30 days of encounter, "no" for no record of readmission

Research Objectives:

- Predict the major factors that are causing the patient to readmit.
- Predict the common age group and the common gender that are getting re-admitted.
- Conduct Exploratory Data Analysis (EDA) on the dataset to form a cluster of common entities.
- Implement K-Means Clustering and Decision Trees to predict the outcome.
- Predict whether a patient will be re-admitted into the hospital.
- Predict whether a patient will be re-admitted into the hospital with-in 30 days of his/her discharge.
- Predict whether a patient will be re-admitted into the hospital after 30 days of his/her discharge.
- Mean age of patients getting re-admitted.
- Average time spent by patients in the hospital before and after re-admission.