

## Project Summary

Batch details	PGPDSE-FT Chennai Sep'21
Team members	1. UMAIR AHMED E F 2. TEJA LAKKADASU 3. VISHAL KARTHICK R 4. MANNE VASU RAJENDRA 5. YOGESHWARAN V
Domain of Project	Healthcare and Life Sciences
Proposed project title	Prediction of Diabetic Patient Readmission
Group Number	1
Team Leader	Umair Ahmed E F
Mentor Name	Mr. Srikar Muppidi

**Dataset name:** diabetic\_data.csv

**Introduction to the problem/domain/background details:** The domain of this Capstone Project falls under the purview of “Healthcare Analytics”. The Capstone project will present the analysis of a large clinical database that was undertaken to examine the historical patterns of diabetes care in patients with diabetes admitted to a US hospital and to indicate future directions which will lead to a reduction in hospital readmission rate and improvements in patient care. The need for readmission indicates that inadequate care was provided to the patient at the time of first admission. Inadequate care poses threat to patients’ life and treatment of readmitted patients leads to increased healthcare costs.

**Problem Statement:** Predicting whether the patient is getting readmitted in the hospital.

**Business problem/ Impact in business of your problem/Need for this study/Abstract (Executive summary):** Patients facing a high risk of readmission need to be identified at the time of being discharged from the hospital, to facilitate improved treatment to reduce the chances of their readmission. Readmission of patients within 30 days of being discharged (short-term readmission) and those with more than 30 days of readmission has been a widely used metric for studying re-admissions. This project aims to carry out the use cases of “Binomial classification” by combining both types of readmissions for research purposes to check the efficiencies of the different ML algorithms.

**Variable identification:*****Independent Variables:***

Encounter ID , Patient Number , Race , Gender , Age , Weight , Admission Type , Discharge disposition , Admission source , Time in hospital , Payer Code , Medical Speciality , Number of Outpatient visits , Number of lab procedures , Number of procedures , Number of Medications , Number of emergency visits , Number of inpatient visits , Diagnosis 1 , Diagnosis 2 , Diagnosis 3 , Number of Diagnoses , Glucose serum test result , A1c test result , Change of medications , Diabetics medication , 24 features for medication.

***Target Variable:*** Readmitted

**Variable information/Data description:**

S.no	Feature name	Description
1.	Encounter ID	Unique identifier of an encounter
2.	Patient Number	Unique identifier of a patient
3.	Race	Values: Caucasian, Asian, African American, Hispanic, and other
4.	Gender	Values: male, female, and unknown/invalid
5.	Age	Grouped in 10-year intervals: [0, 10), [10, 20), ..., [90, 100)
6.	Weight	Weight in pounds
7.	Admission Type	Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, new-born, and not available
8.	Discharge disposition	Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available
9.	Admission source	Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital
10.	Time in hospital	Integer number of days between admission and discharge
11.	Payer Code	Integer identifier corresponding to 23 distinct values, for example, Blue Cross\Blue Shield, Medicare, and self-pay
12.	Medical Speciality	Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct

		values, for example, cardiology, internal medicine, family\general practice, and surgeon
<b>13.</b>	Number of Outpatient visits	Number of outpatient visits in the year preceding the encounter
<b>14.</b>	Number of lab procedures	Number of lab tests performed during encounter
<b>15.</b>	Number of procedures	Number of procedures (other than lab tests) performed during the encounter
<b>16.</b>	Number of Medications	Number of distinct generic names administered during the encounter
<b>17.</b>	Number of emergency visits	Number of emergency visits of the patient in the year preceding the encounter
<b>18.</b>	Number of inpatient visits	Number of inpatient visits of the patient in the year preceding the encounter
<b>19.</b>	Diagnosis 1	The primary diagnosis (coded as first three digits of ICD9) 848 distinct values
<b>20.</b>	Diagnosis 2	Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values
<b>21.</b>	Diagnosis 3	Additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct
<b>22.</b>	Number of Diagnoses	Number of diagnoses entered in the system
<b>23.</b>	Glucose serum test result	Indicates the range of the result or if the test was not taken. Values: ">200," ">300," "normal," and "none" if not measured

<b>24.</b>	A1c test result	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
<b>25.</b>	Change of medications	Indicates if there was a change in diabetic medications (either dosage or generic name). Values: "change" and "no change"
<b>26.</b>	Diabetics medication	Indicates if the there was any diabetic medication prescribed. Values: "yes" and "no"
<b>27.</b>	24 features for medication	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	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

**Future Work/Methodology (Details of algorithms):** Understanding the data, handling the null values, encoding categorical attributes, treating outliers if any on numerical columns containing outliers, splitting the data for training and testing. As the business problem statement based on the classification applying all the algorithms like logistic regression, decision tree, random forest and KNN, boosting algorithms and selecting the best model and further evaluating the model using the Confusion matrix, accuracy, ROC-AUC, Classification report.

**Timeline Chart (Weekly plan):** Tentative weekly plan that you will be following. E.g., as shown below.

Week 1	24-12-2021	Understanding the dataset and deep dive into business domain
Week 2 – Week 3	07-01-2022	Performing EDA and various feature engineering techniques on the dataset, Baseline Model Building (Work progress status I)
Week 4	20-01-2022	Interim Presentation and Interim Report
Week 5 – Week 7	17-02-2022	Work Progress Status II (End-End Model Building, Model Tuning, Cross validation, Model Evaluation, Model Selection)
Week 8	18-02-2022	Final model Presentation and final report submission

**References (Data set source/Journals/articles):** Wang, Tong, Cynthia Rudin, Finale Doshi-Velez, Yimin Liu, Erica Klampfl, and Perry MacNeille. 'A Bayesian framework for learning rule sets for interpretable classification.' The Journal of Machine Learning Research 18, no. 1 (2017): 2357-2393.,

Data Set link: <https://archive.ics.uci.edu/ml/machine-learning-databases/00296/>

Journal link: <https://downloads.hindawi.com/journals/bmri/2014/781670.pdf>

**Declaration:** This is to declare that the dataset that we are using for our capstone project does not have any relevant legality associated to it and can be used to showcase the work we do on it as a presentation in Great Learning.