

### **Background:**

Diabetes is among the most prevalent chronic diseases in the United States. It can lead to reduced quality of life and life expectancy, and can also lead to several complications, including heart disease, vision loss, amputation, and kidney disease. There is no cure for diabetes, but early diagnosis can lead to lifestyle changes and better treatments and outcomes for patients.

### **Problem Statement:**

Our client, Very Fancy Hospital (hereinafter, “VFH”), is the largest hospital network in Northeastern Pennsylvania. VFH is also one of the leading hospitals for diabetes research, and patients from all over the country seek treatment for this chronic disease.

VFH is interested in determining whether it’s possible to predict which of their current patients are likely to become diabetic in the future.

### **Data Sources & Problem Approach:**

This project will use data from Kaggle – <https://www.kaggle.com/alexteboul/diabetes-health-indicators-dataset?resource=download>

The dataset includes 253,680 survey responses to the CDC’s Behavioral Risk Factor Surveillance System (a health-related telephone survey that is collected annually by the CDC). It includes health data from those who either are diabetic, prediabetic, or non-diabetic (or, alternatively, only experienced gestational diabetes).

The business scenario will be modeled by building several classification models, which will be evaluated and compared according to appropriate performance metrics selected according to the goals of the client. Since we are focused on whether a patient is diabetic or prediabetic instead of non-diabetic, it makes sense for us to pursue the use of machine learning classification models towards this problem.

In addition, interpretability analyses will be conducted to characterize how the variation of identified features will affect the probability associated with each of the classes under study.

### **Required Deliverables:**

As required, the following will be delivered at the end of this project: one Jupyter notebook per project phase, a written report, and a presentation slide deck.