**Overview**

The training data was collected and made available by “National Institute of Diabetes and Digestive and Kidney Diseases” as part of the [Pima Indians Diabetes Database](https://www.kaggle.com/uciml/pima-indians-diabetes-database/data). Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here belong to the Pima Indian heritage (subgroup of Native Americans).

We use this Data Set for the initial training purpose. We do improvise the Data Set with values obtained from new customers and all other possible sources.

# Data Description

The following features have been provided to help us predict whether a person is diabetic or not:

* **Pregnancies:**Number of times pregnant. In case of Male, zero by default
* **Glucose:** Plasma glucose concentration over 2 hours in an oral glucose tolerance test
* **BloodPressure:**Diastolic blood pressure (mm Hg)
* **SkinThickness:** Triceps skin fold thickness (mm)
* **Insulin:** 2-Hour serum insulin (mu U/ml)
* **BMI:** Body mass index (weight in kg/(height in m)2)
* **Age:** Age (years)
* **Outcome:** Class variable (0 if non-diabetic, 1 if diabetic)

**Method Followed**

We use a set of **Classification Algorithms**. We have used **Boosting** and **Bagging** methods for greater accuracy.

Our code keeps updating the Data Base and trains all the ML Algorithms to obtain accuracy Values.

It then selects the Algorithm giving maximum accuracy and gives us the prediction. We then request the customer for their feedback about the outcome. Then we save all the inputs and outcome given by the customer into the data base