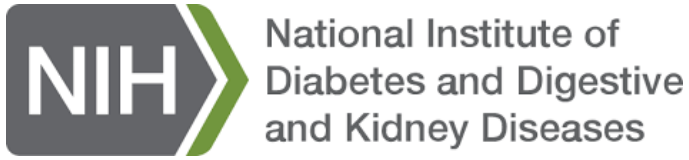
Pima Indians Diabetes Database

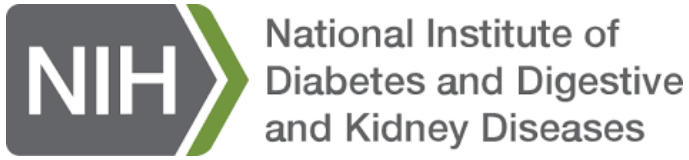
Mini Project 3

Context



- This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases.
- The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.
- Several constraints were placed on the selection of these instances from a larger database.
- In particular, all patients here are females at least 21 years old of Pima Indian heritage.

Details about the dataset:



The dataset consists of several medical predictor variables and one target variable, Outcome. Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

- **Pregnancies**: Number of times pregnant
- **Glucose**: Plasma glucose concentration a 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)
- **DiabetesPedigreeFunction**: Diabetes pedigree function
- **Age**: Age (years)
- **Outcome**: Class variable (0 or 1)
- **Number of Observation Units: 768**
- **Variable Number: 9**

Problem Statement

National Institute of Diabetes and Digestive and Kidney Diseases

- Can you build a machine learning model to accurately predict whether or not the patients in the dataset have diabetes or not?

All The Best