# Introduction:

We will use Naive Bayes model on the \*\***Prima Indians Diabetes**\*\* data set. The model will predict which people are likely to develop diabetes. In particular, all patients here are females at least 21 years old of Pima Indian heritage. The datasets consists of several medical predictor variables.

# Dataset:

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.

# Dataset Description:

Several constraints were placed on the selection of these instances from a larger database. The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables includes 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)

## Steps Involved

1. Importing the libraries

2. Data Import

3. Data Preprocessing

4. Study Correlation

5. Train and test data split

6. Build model

7. Make predictions

## Machine Learning Model used:

1. Naive Bayes

## Libraries Involved:

1. Numpy

2. pandas

3. matplotlib

4. sklearn

5. seaborn

## Visualizations:

1. histogram

2. correlation

3. pairplot

4. confusion matrix