**Logistic Regression Assignment**

**Pima Indians Diabetes Database**

Machine learning techniques are gaining an increasing significance in the field of medical research especially in the diagnosis of Diabetes. It is estimated that about 415 million people are suffering from this deadly disease, for every 1 in 11 of the world’s adult population.

This dataset which was originally documented by the National Institute of Diabetes and Digestive and Kidney Diseases, consists of patients of Pima Indian Heritage, exclusively females, above 21 years of age.

**Attributes of diabetes.csv:**

Pregnancies: Number of times pregnant

Glucose: Plasma glucose concentration in an oral glucose tolerance test

Blood Pressure: Diastolic blood pressure (mm Hg)

Skin Thickness: Triceps skin fold thickness (mm)

Insulin: 2-Hour serum insulin (mu U/ml)

BMI: Body mass index (weight in kg/(height in m)^2)

Diabetes Pedigree Function: Diabetes pedigree function, data on diabetes occurrence history in relatives and the genetic relationship of those relatives to the patient

Age: Age in years

Outcome: Class variable (0 or 1)

**File description:**

diabetes.csv-dataset describing the above mentioned attributes.

**Problem:**

The objective is to predict based on available data whether a patient has diabetes or not with the help of logistic regression.