Diabetes Prediction

# Abstract:

Diabetes mellitus (DM), commonly referred to as diabetes, is a group of metabolic disorders in which there are high blood sugar levels over a prolonged period. Type 1 diabetes results from the pancreas's failure to produce enough insulin. Type 2 diabetes begins with insulin resistance, a condition in which cells fail to respond to insulin properly. As of 2015, an estimated 415 million people had diabetes worldwide, with type 2 diabetes making up about 90% of the cases. This represents 8.3% of the adult population

# Problem Statement:

The objective is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.

# Dataset Information:

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. 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.

# Variable Description:

|  |  |
| --- | --- |
| **Column** | **Description** |
| Pregnancies | Number of times pregnant |
| Glucose | Plasma glucose concentration a 2 hours in an oral glucose tolerance test |
| BloodPressure | Diastolic blood pressure |
| SkinThickness | Triceps skin fold thickness |
| Insulin | 2-Hour serum insulin |
| BMI | Body mass index |

|  |  |
| --- | --- |
| DiabetesPedigreeFunction | Diabetes pedigree function |
| Age | Age in years |
| Outcome | has diabetes or not |

**Scope:**

* Check missing values and outliers and treat them accordingly
* Feature Selection and Data Pre-processing
* Evaluating the model with various metrics like Accuracy, AUC ROC, Precision, etc. and improve the score using statistical analysis over time