**Data and AI Training for Ext LT-PHDSI**

**Classic ML – Diabetic Prediction Model**

## 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.

## Content

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, Glucose, SkinThickness and so on.

## Problem Statement

Predict the onset of diabetes based on diagnostic measures.

As part of the Exercise – following concepts will be explored

* Loading the Dataset from CSV file to Data frame using Python – pandas lib
* Exploratory data analysis : EDA
* Data visualization and interpretation
* Data Imputation
* Feature co-relation and heat map
* Data scaling
* Model Building
* Model Performance
* Model Saving
* Prediction with some random data

Dataset will be shared as part of the GitHub repo - <https://github.com/vallerivinodh/AI-Examples>

## Tools Used

Jupyter Notebook with Online tool – [Google Colab](https://colab.research.google.com/) - <https://colab.research.google.com/>

Python 3.x