

**S**PECIFIC

* We are a group of three and our project is based on Multivariate Analysis.
* This dataset we have selected is from National Institute of Diabetes and Digestive and Kidney Diseases.
* The aim of our project is to diagnose if a particular patient has diabetes based on various medical parameters in the dataset.
* We will accomplish this by using machine learning techniques

Technologies used: R and Rstudio

**M**EASURABLE

* Perform operations on the dataset depending upon the nature of the dataset and the questions asked.
* The performance can be measured based on the completion of the assigned task and also on the results from the techniques performed.

**A**CHIEVABLE

* Operations performed on the pima diabetes dataset:
* Understood the nature of dataset.
* Asked relevant questions.
* Cleaned the data
* Data Visualization
* Performed Statistical tests

**R**ELEVANT

* The power of machine learning in diagnosing disease and in sorting and classifying health data will empower physicians and speed-up decision making in the clinic.
* Using different analytic techniques and algorithms can provide better information to doctors at the point of patient care.

**T**IME-BOUND

* The long term goal is to predict if a patient has diabetes or not based on a few medical factors.
* Research and perform the best algorithms which will help us in accurate prediction.



* Here, the ‘Outcome’ variable is the Dependent variable. This variable depends on the variation of the other independent variables.
* Each independent variable has a different impact on the output variable.



The following are the independent variables

* Pregnancies – The number of times the women is pregnant
* GlucosePlasma - Glucose concentration in an oral glucose tolerance test
* BloodPressure- Diastolic blood pressure (mm Hg)
* SkinThicknessTriceps -skin fold thickness (mm) #Insulin2- Hour serum insulin (mu U/ml)
* BMIBody mass index- (weight in kg/(height in m)^2)
* DiabetesPedigreeFunctionDiabetes -pedigree function
* #Age -(years)
* #OutcomeClass variable – Dependent Variable

All are numeric variables except for the outcome variable (yes/no)

