**Roshith K R** - RVCE20MCA009

**Akshay T N** – RVCE20MCA069

**Phase 1 Machine Learning:**

1) **Problem Definition:**

Diabetes is a Chronic disease characterized by hyperglycemia. It is characterized by high blood glucose levels resulting from defects in insulin production, insulin action or may be both. based on their blood pressure, and other we will predict which is more beneficial/important to get the accurate output.

our dataset consists of pregnancies woman getting infected to diabetes along with glucose, insulin level, skin-thickness and BMI. we have further classified into age and diabetes pedigree function to get the accurate result he/she is diabetic.

2) **What, Why and How:**

**What** – assuming that we are a disease prediction center, on what particular aspect the

person get infected to disease will find out here, so that we will get the outcomes of the

positive diabetic patient.

**Why –** solving this problem will help to sort or figure out which persons will get disease

diabetes in what stage or reason.

**How –** we can solve this issue/medical problem using previous set of medical data, and

using an KNN-Algorithm.

3) **Data Set and Features Considered:**

**Data Set** – Diabetic Prediction Data set scraped from National Institute of Diabetes.

**Features Considered** - Glucose, blood-sugar, Age, BMI, Diabetes Pedigree function,

Thickness, Pregnancies, Skin-thickness and outcome.

Dataset has about 769 Patient Records.

4) **Type of Algorithm to be used and Why!?:**

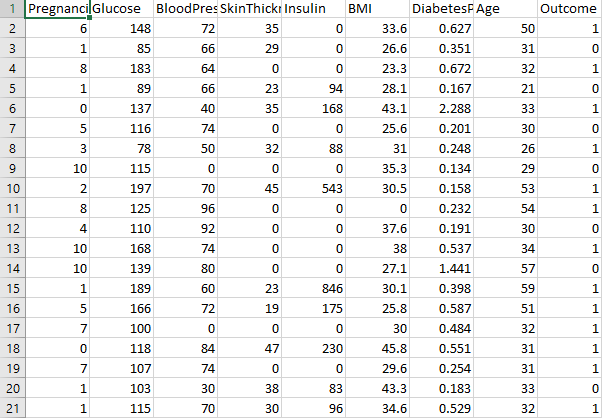
**Type** - K-Nearest Algorithm (KNN) Based on Supervised Learning Technique.

**Why –** Since we will be predicting diabetic or not, we will use KNN because it uses data

to build a model and then uses that model to predict the outcome in 0 or 1 and which

makes the disease risk prediction model more effective than other classifiers.

5) **Sample Dataset With Explanation:**

****

**NOTE:**

**Glucose:** The Glucose level of the person and it is the main sugar found in our blood of the person.

**Blood-pressure:** The blood pressure range or level of the person which may be high or low in the person.

**Age:** The ages of a particular person will get infected. here, the particular person age is available.

**BMI:** BMI will give a body mass index to measure the body fat mass of the particular person.

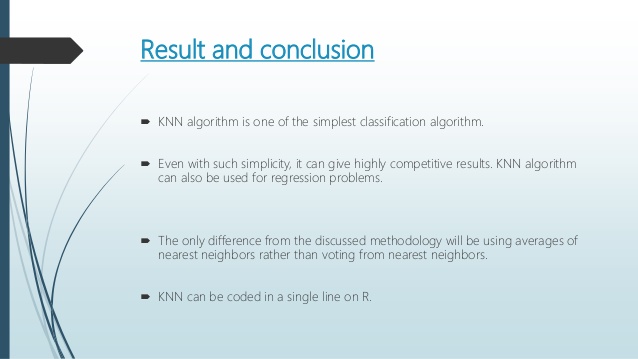
**Diabetes Pedigree Function:** It is a function which gives scores/values based on the family history of whether a patient is diabetic or not in bit like 0 or 1.

**Skin Thickness:** It gives thickness of the skin when the person gets diabetic. Automatically skin get thickness.

**Pregnancies:** The number of pregnancy woman getting high blood sugar levels in pregnancy time and also chances of getting infected to diabetes. here, number of persons data given.

**Outcome:** It will give the final result as whether the person is diabetic or not in bit values like 0 or 1. If 0 means he/she is non diabetic and 1 means he/she is diabetic. It is a target variable.

6) **KNN-Algorithm:**

****