PROBLEM STATEMENT

Before starting with the project and using some machine learning stuff to solve problem, it is very important to know about the importance of each variable which represents each characteristic that helps inturns to achieve our objective (or its Garbage In and Garbage Out).

Our approach for solving the problem will be answering to three questions mainly 1.What is the Objective? 2. Why it is important to achieve objectve? 3. How to achieve Objective?

What is the Objective ?

The objective of this is to identify which treatment is effective among Solo Insulin or Insulin Combined with other drugs for the diabetic patients and building a predictive model to predict effective treatment for a diabetic patient.

Why it is important to achieve objective?

Since,From many years people are suffering from the diabeties but the diabeties is only controlled when right medication is given to the patients identifying the charateristics of a patients and treating accordingly.

How to achieve the objective ?

For this analysis the data is collected from 130 US hospitals over 10 years(1999-2008).It includes over 50 features representing patient and hospital outcomes.Here one can find the solution to this complex problems by using Machine Learnig.

Let's start our analysis by understanding each and every input variable .....

From the dataset the following attributes are used for identifying unique patinets.

Patinent_nbr : This attribute is used to identify unique patientsadmitted into the hospitals.

Encounter_id : This attribute gives information of Unique identifier of an patient encounter.