

# AI Based Diabetes Prediction System

## Problem Definition:

The problem is to build an AI-powered diabetes prediction system that uses machine learning algorithms to analyze the medical data and predict the likelihood of an individual developing diabetes. The system aims to provide early risk assessment and personalized preventive measures, allowing individuals to take proactive actions to manage their Health.

## Preview:

1. Collect the medical dataset
2. After that pre-processing the data
3. Analyze the dataset
4. Provide the accurate result.

## Data Collection:

The dataset must contain the medical feature such as **glucose level, blood pressure, BMI**, etc., along with information about whether the individual has diabetes or not. The dataset are obtained from Kaggle.com.

## Dataset:

B	C	D	E	F	G	H	I
Glucose	Blood pressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
148	72	35	0	33.6	0.627	50	1
85	66	29	0	26.6	0.351	31	0
183	64	0	0	23.3	0.672	32	1
89	66	23	94	28.1	0.167	21	0
137	40	35	168	43.1	2.288	33	1
116	74	0	0	25.6	0.201	30	0
78	50	32	88	31	0.248	26	1
115	0	0	0	35.3	0.134	29	0
197	70	45	543	30.5	0.158	53	1
125	96	0	0	0	0.232	54	1
110	92	0	0	37.6	0.191	30	0
168	74	0	0	38	0.537	34	1
139	80	0	0	27.1	1.441	57	0
189	60	23	846	30.1	0.398	59	1
166	72	19	175	25.8	0.587	51	1
100	0	0	0	30	0.484	32	1
118	84	47	230	45.8	0.551	31	1
107	74	0	0	29.6	0.254	31	1

### **Parameters Description:**

If a person have BMI level 30 or greater, Blood pressure 130/80 mm hg and glucose level before a meal:80 to 130 mg/dl.Two hours after the start of a meal:less than 180 mg/dl the person have diabetes.

### **Data pre-processing:**

This phase of model handles inconsistent data in order to get more accurate and precies results.If the dataset contains missing value,we imputed missing values for few attributes like Gulcose levels,Blood pressure,BMI,age because these attributes cannot contain zero values.Then we scale the dataset to normalize all values.

### **Features Selection:**

- Using this model we can easily predict the diabetes occur to the person in future or not and prevent from the diabetes.
- It also sent an alert message to the person and suggest some food pattern that are suggest by the Doctors(Diabetes Speacialist).

### **Model Selection:**

In this phase we use various machine learning alorithm for predict the diabetes.These algorithms like **Random forest Classifier,Logistic Regression and Gradient Boosting.**

### **Evaluation:**

We evaluate the prediction results using various evaluation metrices like accuracy,precision,F1-score,recall.

### 1. Accuracy:

It is the ratio of number of correct prediction to the total number of input samples. It is given as

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions made}}$$

### 2. Precision:

It is the number of correct positive results divided by the number of positive results predicted by the classifier. It is expressed as

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

### 3. Recall:

It is the number of correct positive results divided by the number of all relevant samples. It is expressed as

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

### 4. F1-score:

It is used to measure a test's accuracy. F1 score is the Harmonic Mean between precision and recall. The range for F1-score is [0,1]. Mathematically, it is given as

$$F1 = 2 * \frac{1}{(1/\text{precision}) + (1/\text{recall})}$$

### Iterative Improvement:

In this model we use the machine learning algorithm like logistic regression, Random Forest, etc., so it provides the accuracy result.