

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ

**“By the name of God (Allah), the all-Merciful, the especially-Merciful”.**

# Diabetes Prediction using classification method

Presented by:

Irfan Ullah Khan  
(BC19040474)

Supervised by:

Mam Warda Fiaz



Submitted for the partial fulfillment of the requirements for the degree of  
BS Computer Science  
in  
Virtual University of Pakistan

# Table of Content

**1**

Problem Statement & Challenges



**2**

Proposed Methodology

**3**

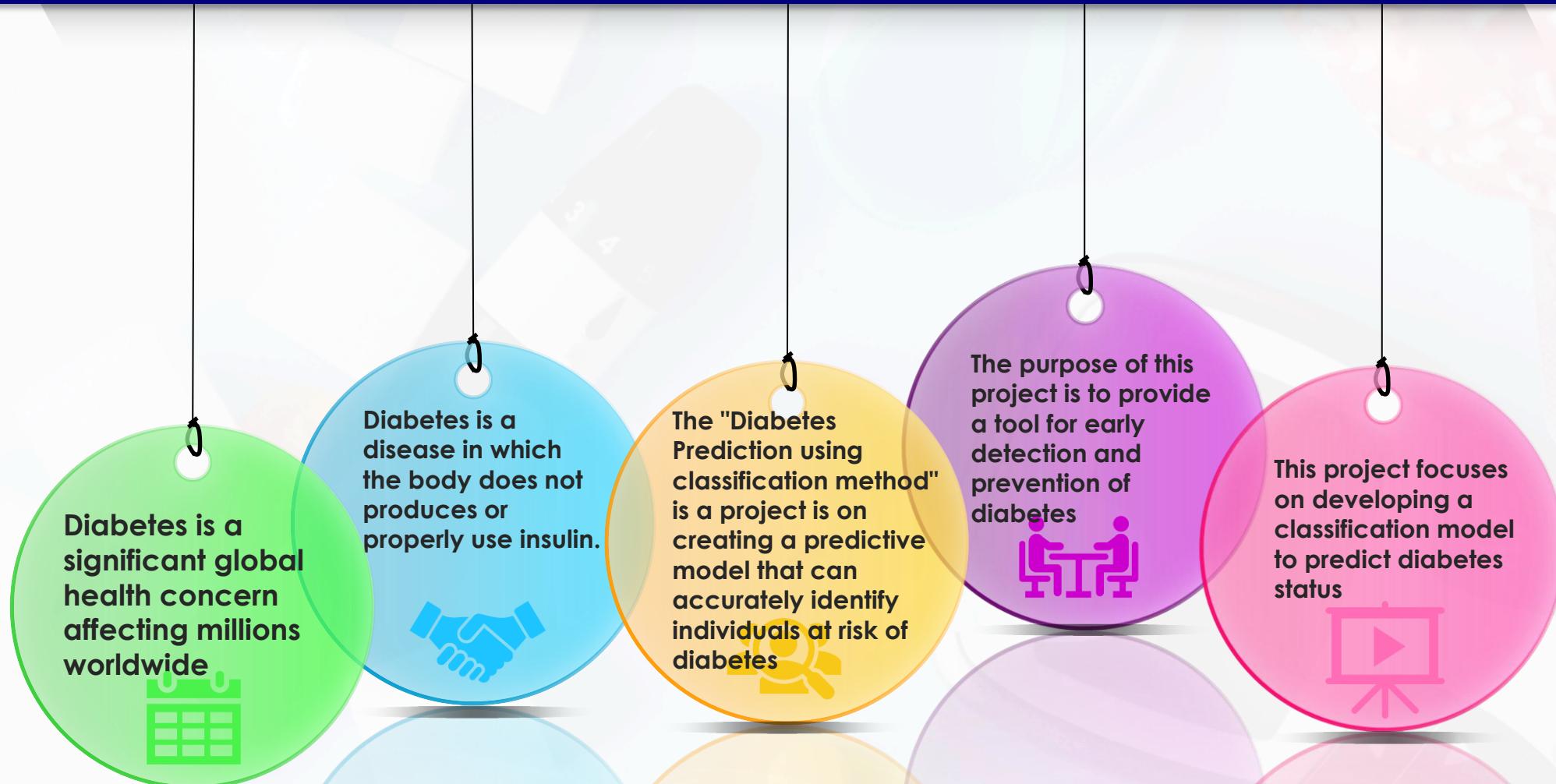
Implementation

**4**

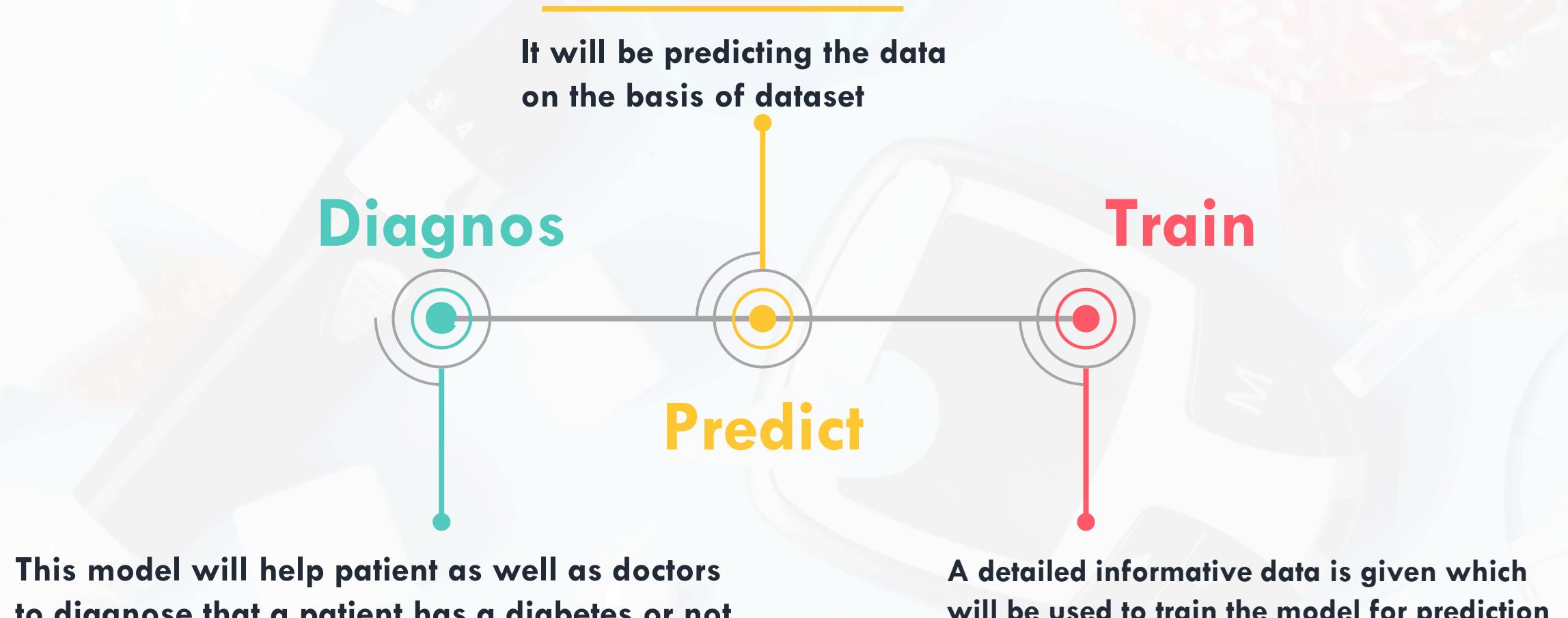
Result & Discussion



# Introduction



# Problem Statement



- 1 • This model will help patient as well as doctors to diagnose that a patient has a diabetes or not.
- 2 • It will be predicting the data on the basis of dataset.
- 3 • A detailed informative data is given which will be used to train the model for prediction.

# Objectives

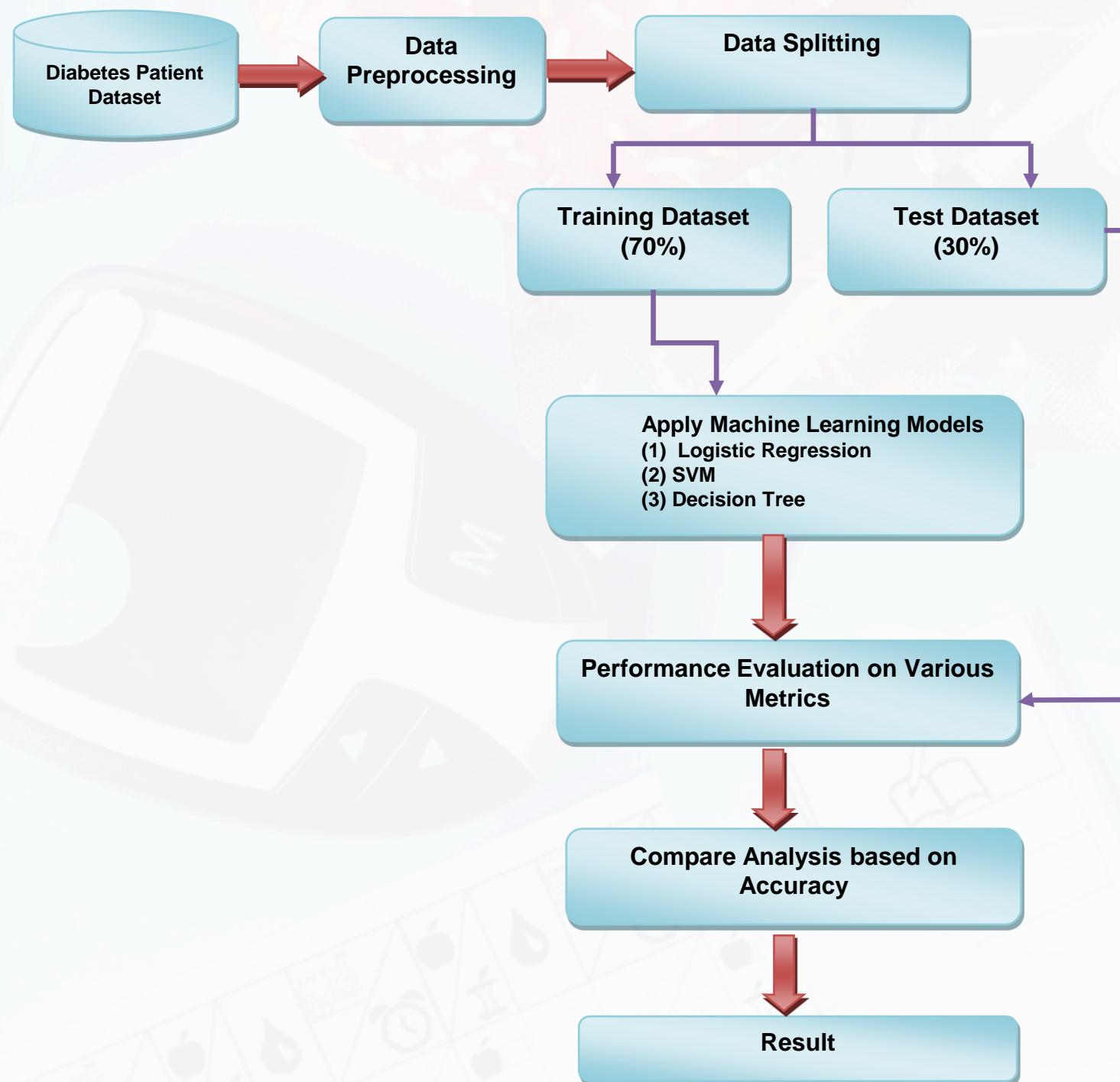
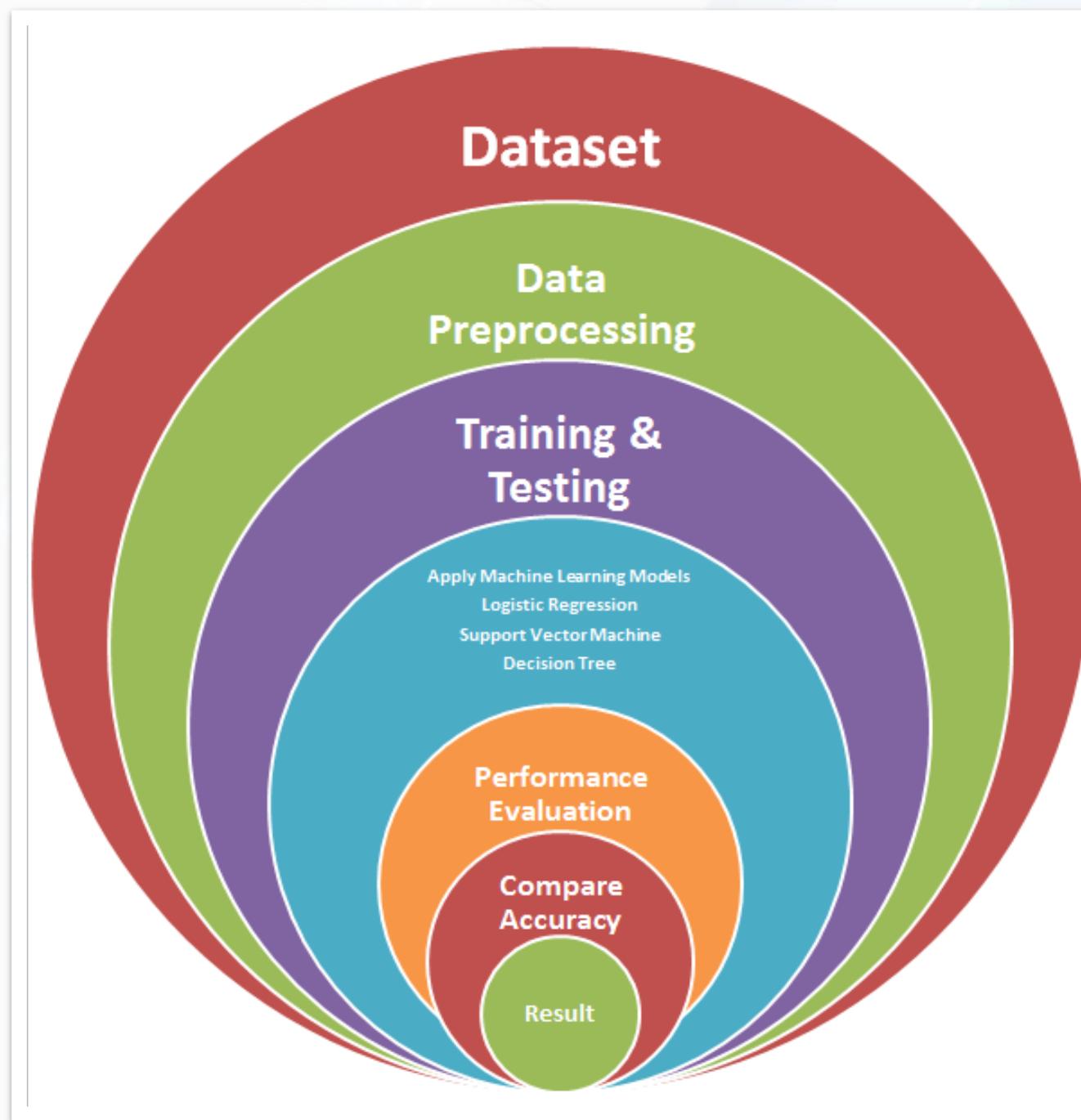
*Identifying hidden patterns and relationships among various attributes that can lead to:*



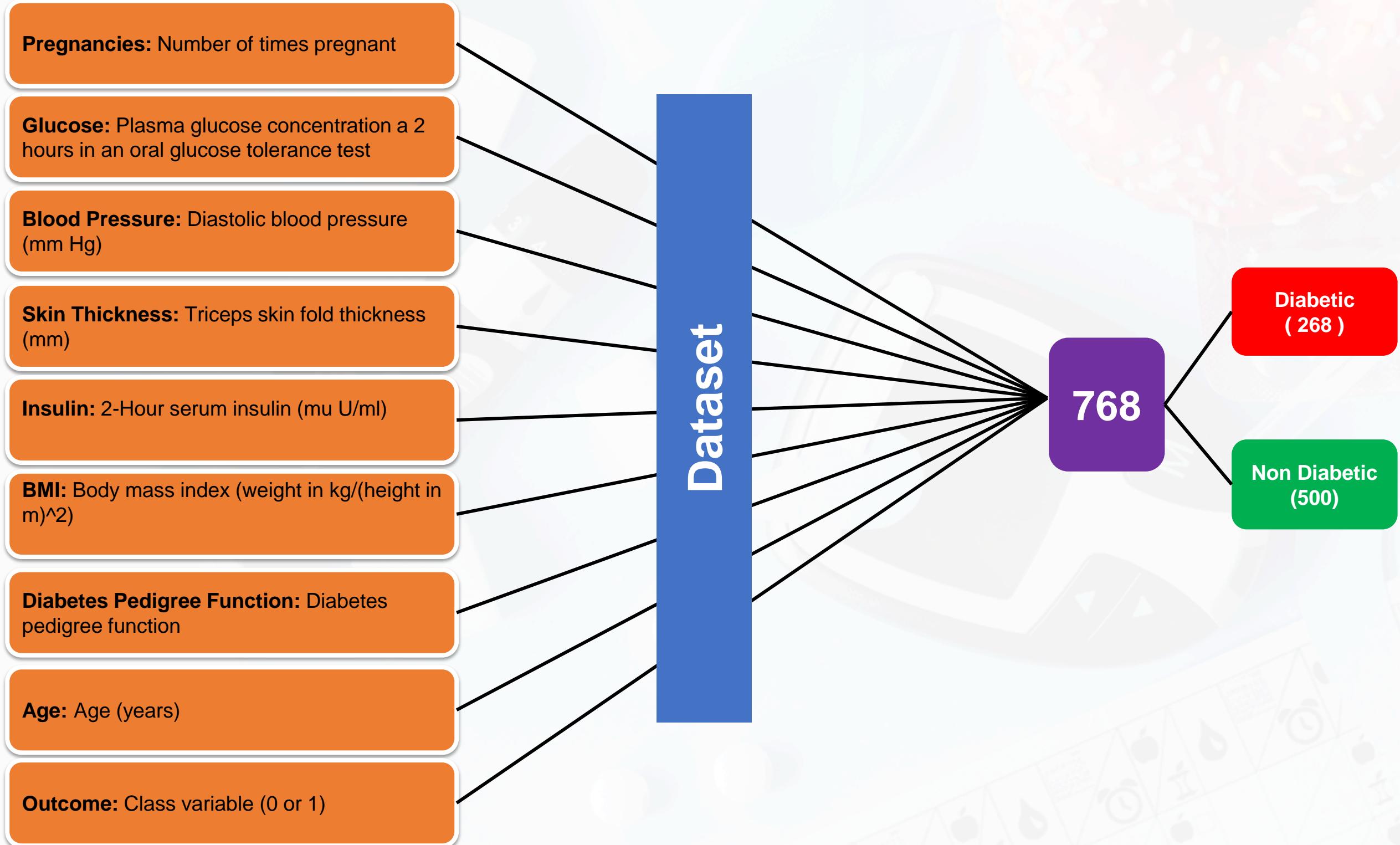
*Early diagnosis may predict the chances of Disease and lead to take preventive measures before the situation becomes critical*

# Proposed Methodology

- ✓ Methodology is a collection of methods, practices, processes, techniques, procedures used for design / execution of your project

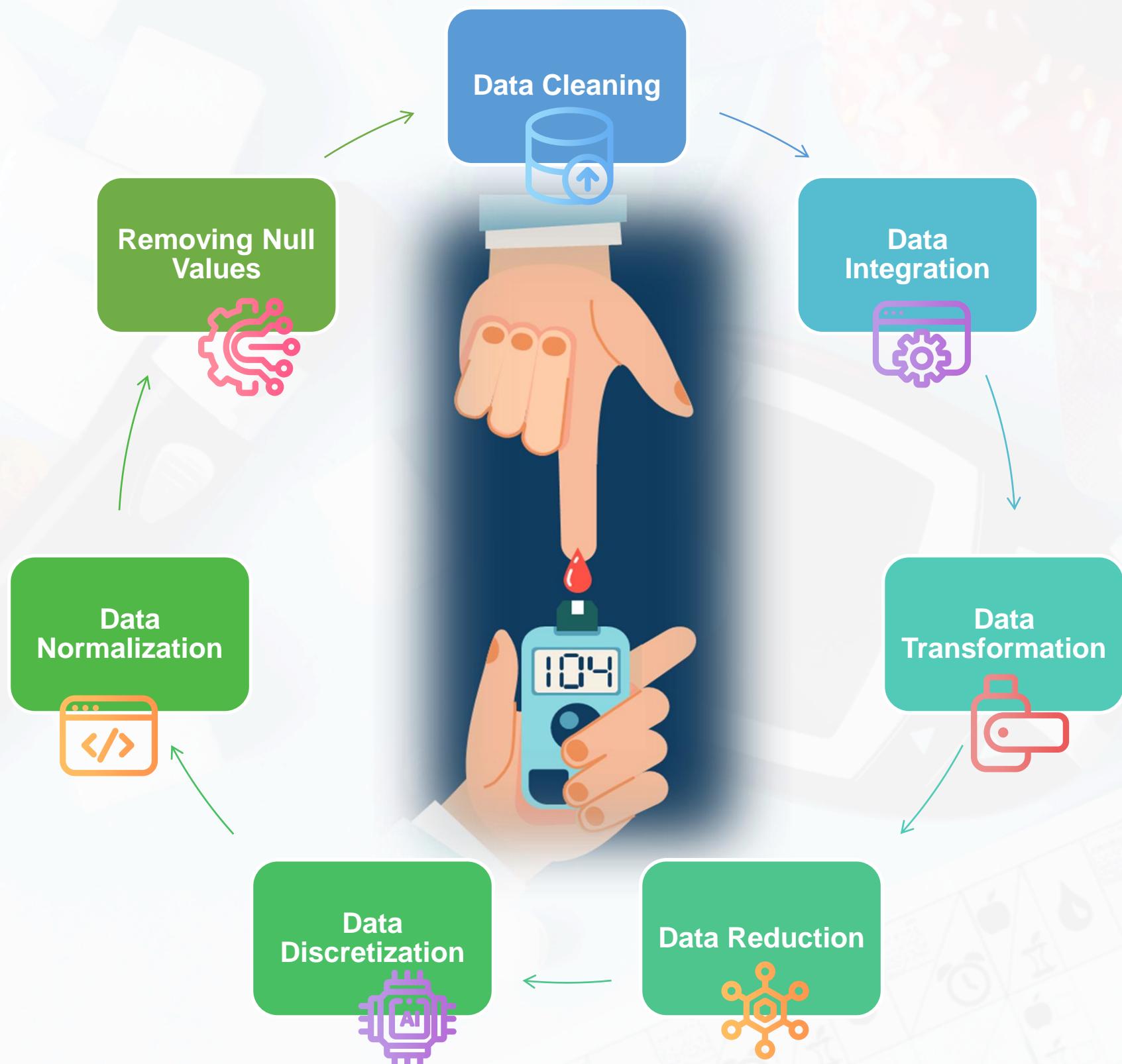


# Description of Dataset



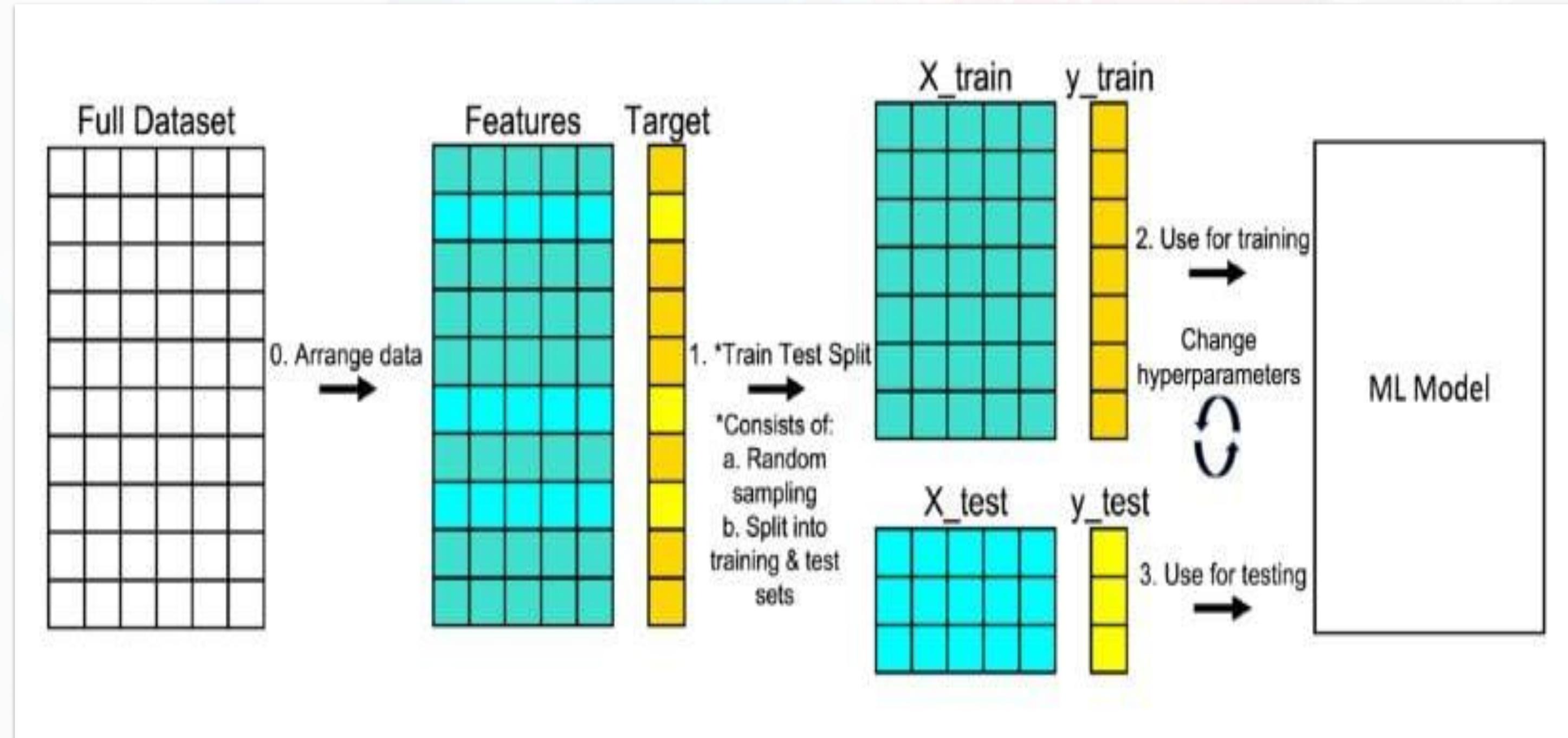
| Index | regnancie | Glucose | bloodPressure | skinThickness | Insulin | BMI  | sPedigree | Age | Outcome |
|-------|-----------|---------|---------------|---------------|---------|------|-----------|-----|---------|
| 0     | 6         | 148     | 72            | 35            | 0       | 33.6 | 0.627     | 50  | 1       |
| 1     | 1         | 85      | 66            | 29            | 0       | 26.6 | 0.351     | 31  | 0       |
| 2     | 8         | 183     | 64            | 0             | 0       | 23.3 | 0.672     | 32  | 1       |
| 3     | 1         | 89      | 66            | 23            | 94      | 28.1 | 0.167     | 21  | 0       |
| 4     | 0         | 137     | 40            | 35            | 168     | 43.1 | 2.288     | 33  | 1       |
| 5     | 5         | 116     | 74            | 0             | 0       | 25.6 | 0.201     | 30  | 0       |
| 6     | 3         | 78      | 50            | 32            | 88      | 31   | 0.248     | 26  | 1       |
| 7     | 10        | 115     | 0             | 0             | 0       | 35.3 | 0.134     | 29  | 0       |
| 8     | 2         | 197     | 70            | 45            | 543     | 30.5 | 0.158     | 53  | 1       |
| 9     | 8         | 125     | 96            | 0             | 0       | 0    | 0.232     | 54  | 1       |
| 10    | 4         | 110     | 92            | 0             | 0       | 37.6 | 0.191     | 30  | 0       |
| 11    | 10        | 168     | 74            | 0             | 0       | 38   | 0.537     | 34  | 1       |
| 12    | 10        | 139     | 80            | 0             | 0       | 27.1 | 1.441     | 57  | 0       |
| 13    | 1         | 189     | 60            | 23            | 846     | 30.1 | 0.398     | 59  | 1       |
| 14    | 5         | 166     | 72            | 19            | 175     | 25.8 | 0.587     | 51  | 1       |
| 15    | 7         | 100     | 0             | 0             | 0       | 30   | 0.484     | 32  | 1       |

# Data Processing

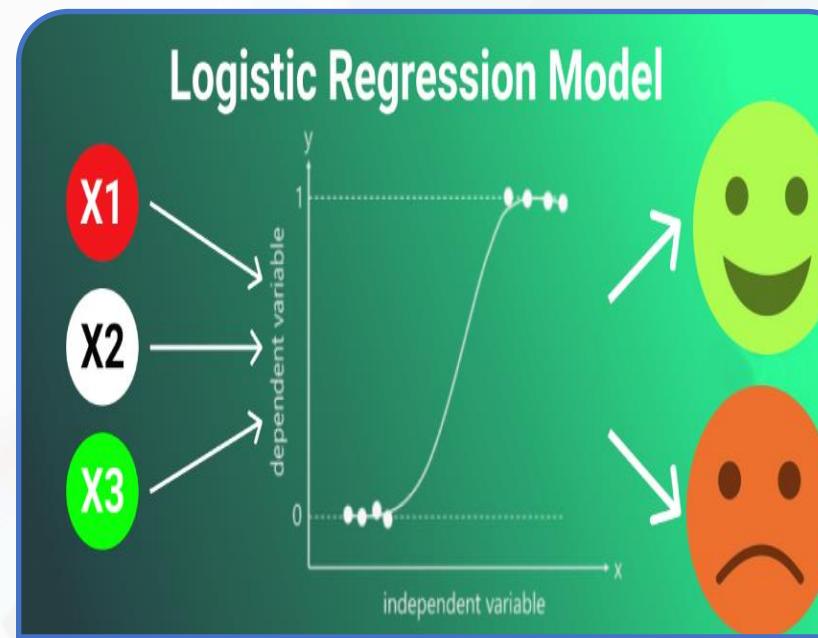


# Train & Test

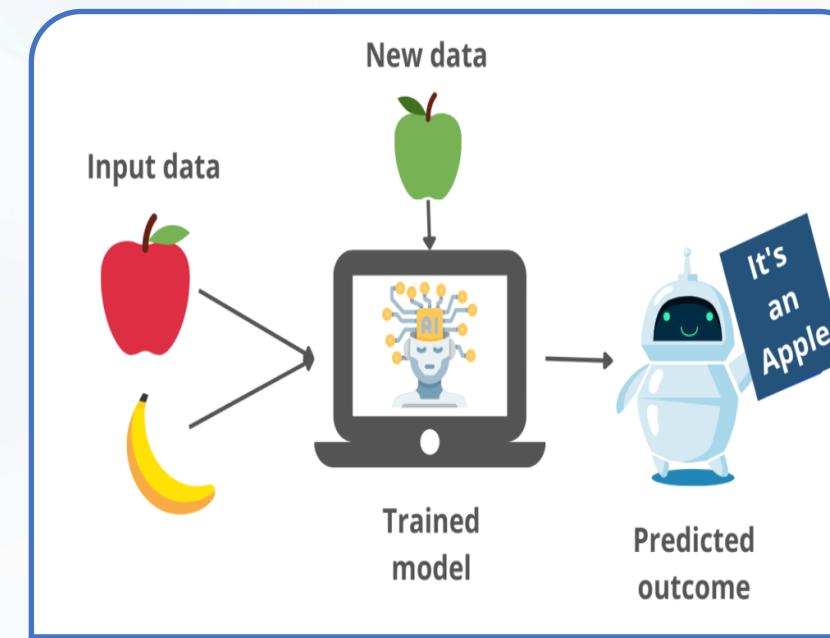
✓ 70% train data & 30% test data.



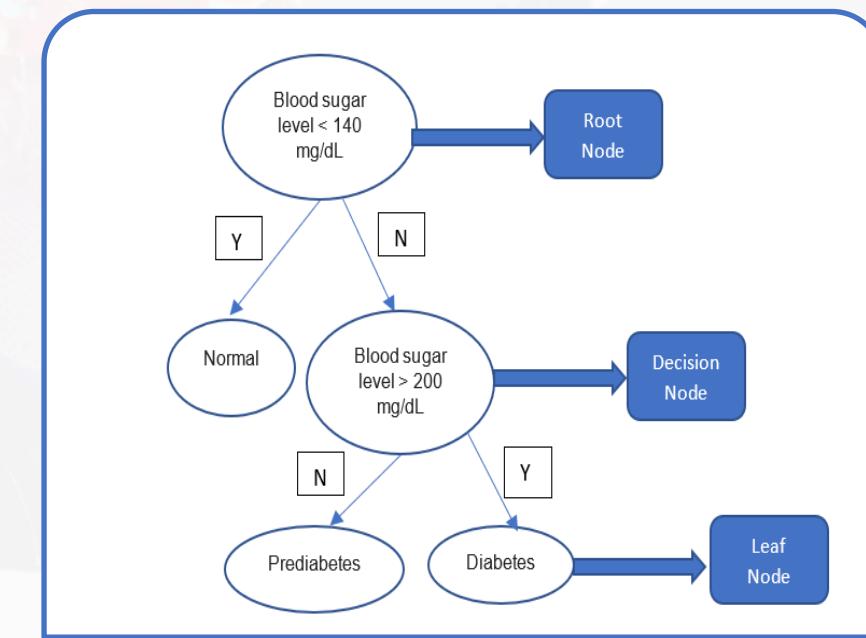
# Algorithm Used in :



**Logistic Regression:**  
Linear model for  
binary classification.

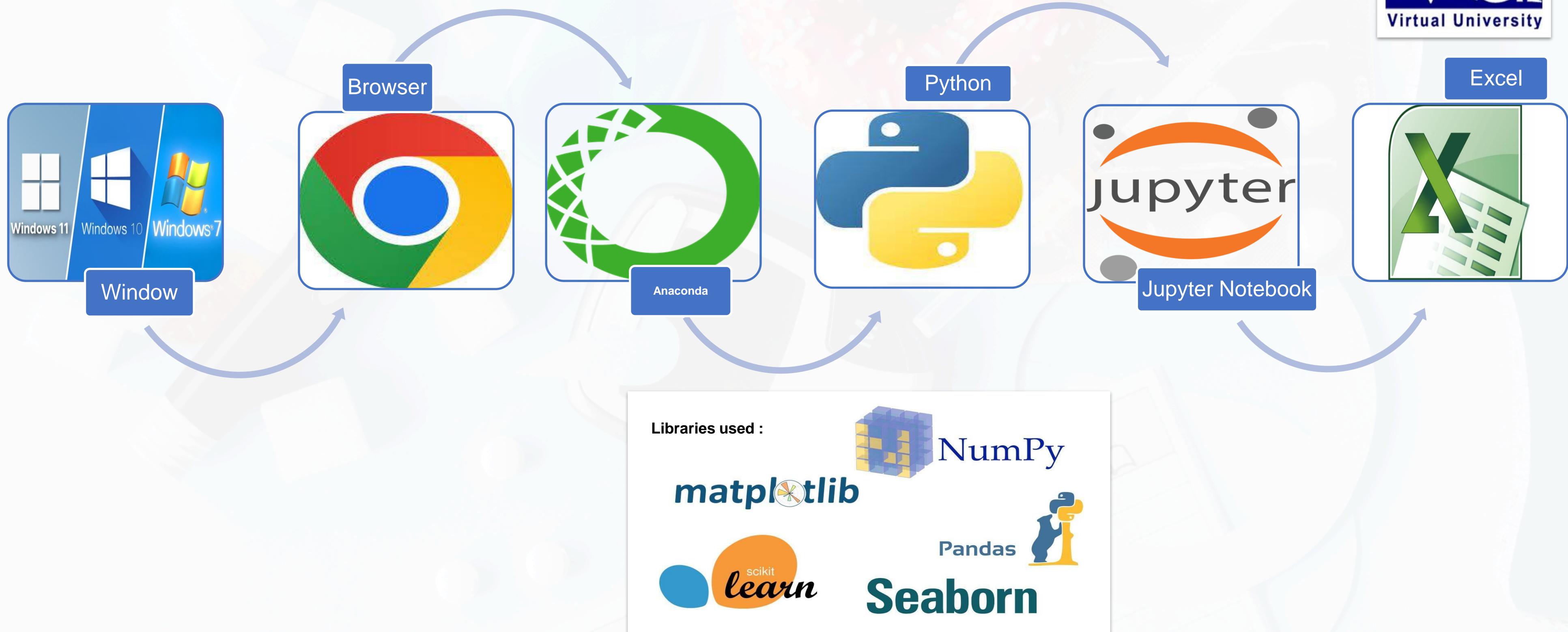


**Support Vector Machine**  
Maximize margin for  
separating classes.



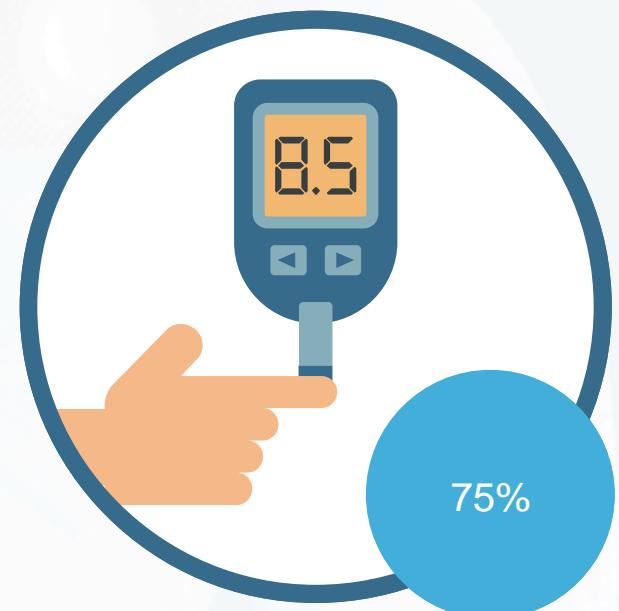
**Decision Tree:**  
Hierarchical tree for  
classification/regression.

# Tools and technologies



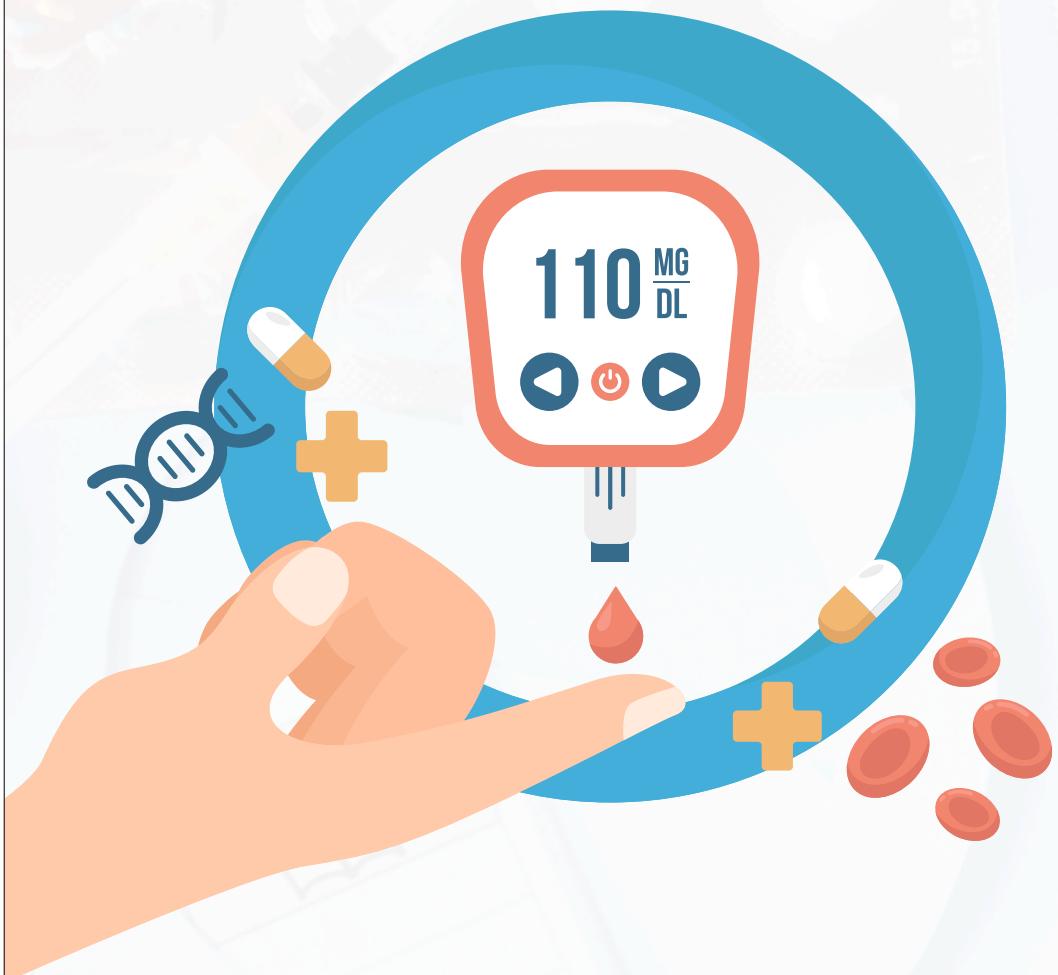
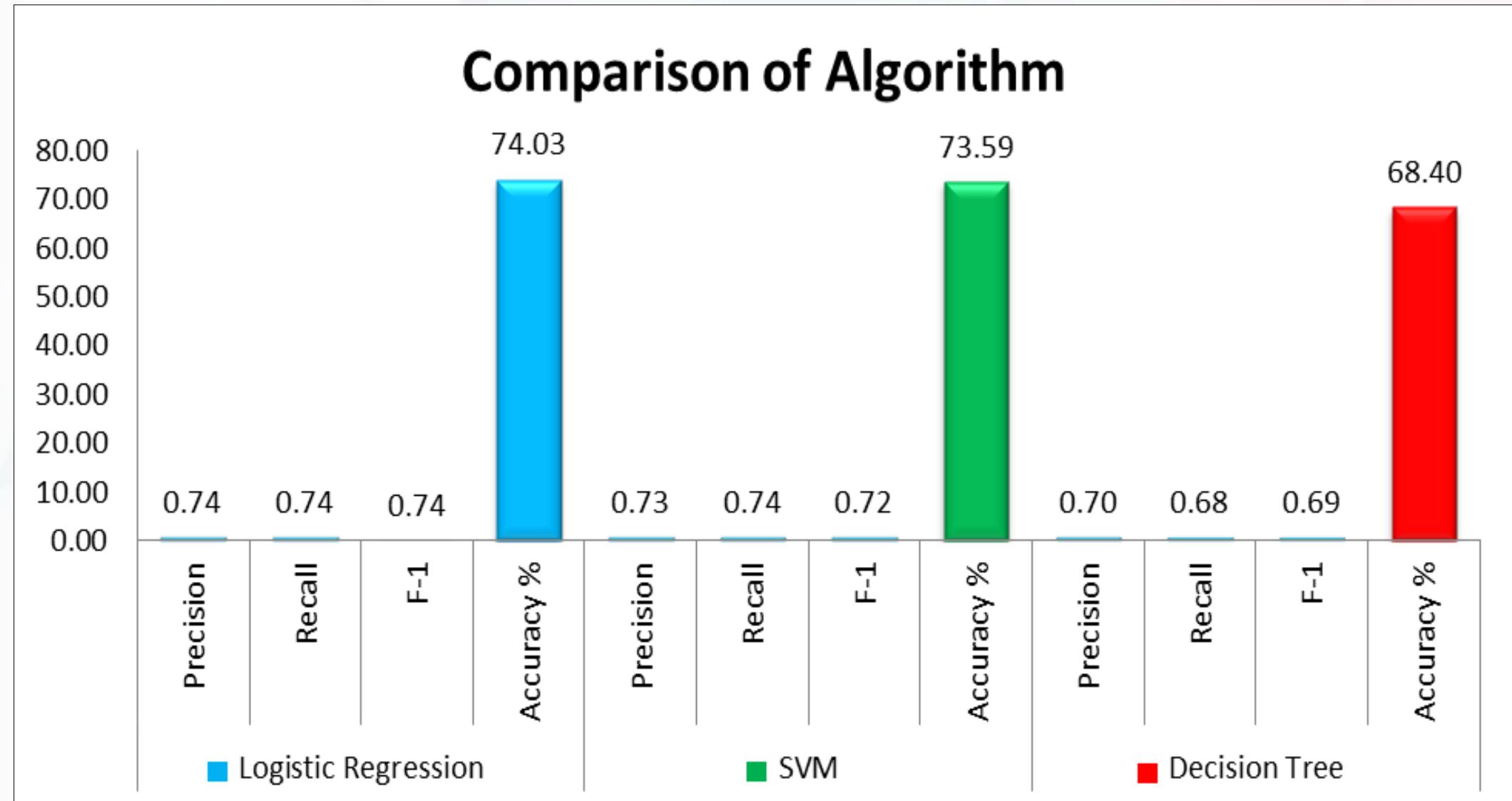
# Discussion For Best

| Algorithm           | Precision | Recall | F-1  | Accuracy %       |
|---------------------|-----------|--------|------|------------------|
| Logistic Regression | 0.74      | 0.74   | 0.74 | <b>74.025974</b> |
| SVM                 | 0.73      | 0.74   | 0.72 | <b>73.593074</b> |
| Decision Tree       | 0.70      | 0.68   | 0.69 | <b>68.396268</b> |



# Comparison of Algorithm (Diagram)

✓ So the best model is “Logistic Regression” which gives 74.03 % Accuracy.



A pile of colorful sticky notes scattered across a surface. Each note is a different color (yellow, pink, blue, green) and features the words "thank you" written in a black, hand-drawn style font. The notes overlap each other, creating a dense, layered effect.