**AI Based Diabetes Prediction System.**

Loading and preprocessing the dataset: Phase\_3

**1 .Download the dataset:**

Download the kaggle dataset given using the link <https://www.kaggle.com/datasets/mathchi/diabetes-data-set>.

**2 .Data exploration:**

After downloading the dataset from kaggle.The dataset format is in the format .csv file.

Start exploring the dataset using the python libraries like pandas.



**3 .Data preprocessing:**

Perform the data-preprocessing using the previously explored data-set.

Preprocessing includes the handling of missing values, feature engineering and data scaling.

\* Missing values-

Check for missing values and decide whether to impute them or remove rows/columns with missing data.



\* Feature selection-

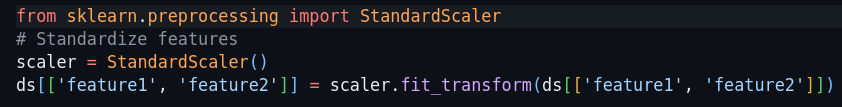
Identify the features that are relevant for diabetes prediction. This might require domain knowledge or statistical analysis.

You can use techniques like correlation analysis to determine feature importance.



\* Feature engineering-

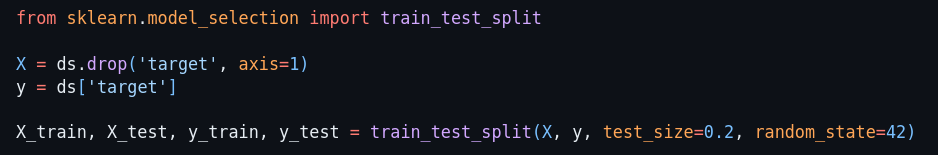
Create new features or transform existing ones that might be useful for the prediction task. This can involve mathematical operations or encoding categorical variables.



**4 .Data splitting:**

Slit the data into training and testing sets for model evaluation .

The typical split ratio is 70-80% for training & 20-30% for testing.



--------------------------------------------------------------------