**TEAM 16: ARTIFICIAL INTELLIGENCE**

**PROJECT TITLE:** **DIABETES PREDICTION SYSTEM**

**PHASE IV: Development part 2**

**DESCRIPTION:**

A diabetes prediction system is a technology or model that uses data analysis and artificial intelligence (AI) techniques to predict the risk of developing diabetes in individuals or to help manage the condition in those who are already diagnosed. Such systems are valuable for early intervention and personalized care. Here's an overview of how a diabetes prediction system typically works in below

**Platform used for implementing:**

We have to use **Jupyter Notebook** for implementing the Machine Learning Algorithm. It have default libraries installed in it. It basically run on a Python program.

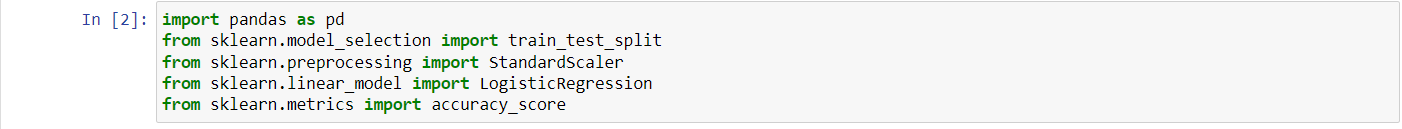
**Sample Program:**

1. **Import libraries:**

**In Python, you use the import keyword to make code in one module available in another. Imports in Python are important for structuring your code effectively.**

Some important libraries known as in python is **Pandas, NumPy**

* **import pandas as pd**
* **from sklearn.model\_selection import train\_test\_split**
* **from sklearn.preprocessing import StandardScaler**
* **from sklearn.linear\_model import LogisticRegression**
* **from sklearn.metrics import accuracy\_score, precision\_score, recall\_score**
* **from sklearn.metrics import f1\_score, roc\_auc\_score, confusion\_matrix**

****

1. **Dataset:**

* Dataset is a collection of various types of data stored in a digital format
* Data is the key component of any Machine Learning project. Datasets primarily consist of images, texts, audio, videos, numerical data points, etc., for solving various Artificial Intelligence challenges such as. Image or video classification.
* Dataset contains the attributes, attributes in which it input from the user and gives output
* Basically the dataset in file format especially in CSV format.

We have to load the dataset into program by in a **csv format**

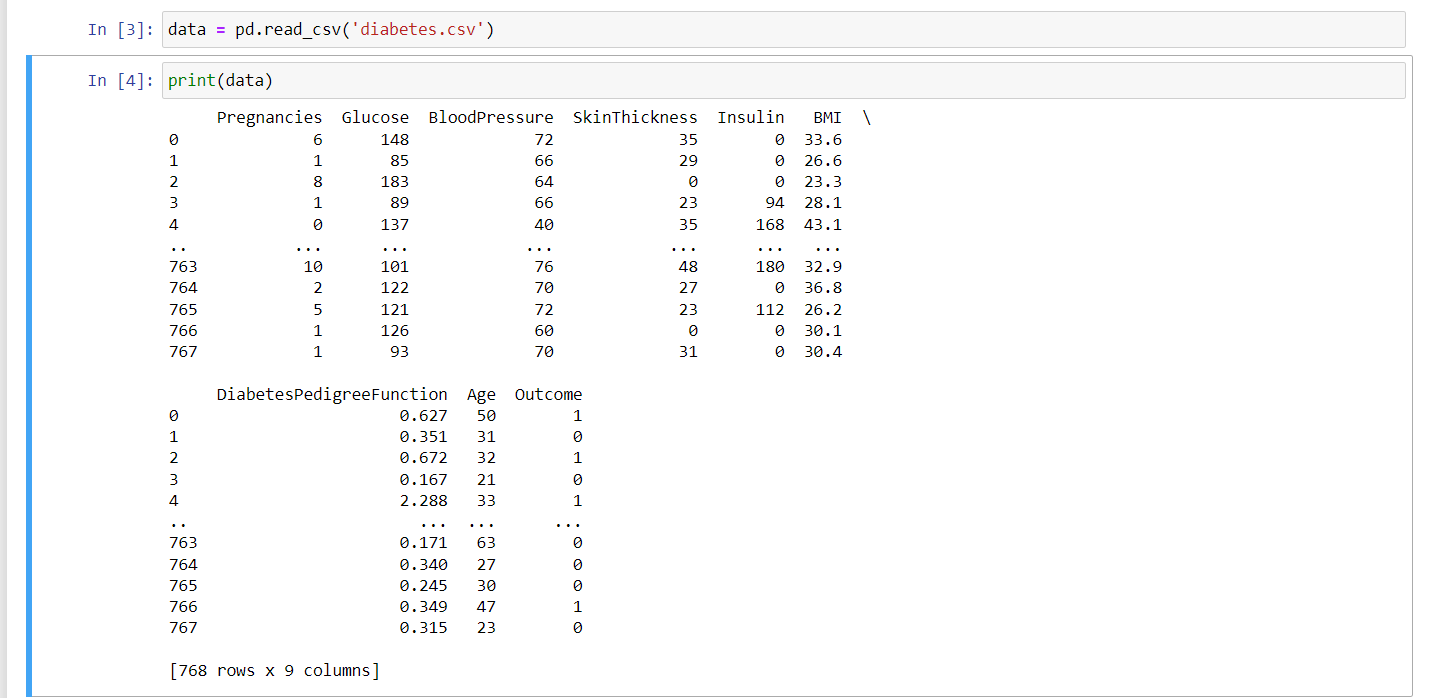
**Dataset Link:**[**https://www.kaggle.com/datasets/mathchi/diabetes-data-set**](https://www.kaggle.com/datasets/mathchi/diabetes-data-set)

* **Loading the dataset into a program by**

**data = pd.read\_csv("diabetes.csv")**

In our Diabetes prediction system dataset contains attributes such as

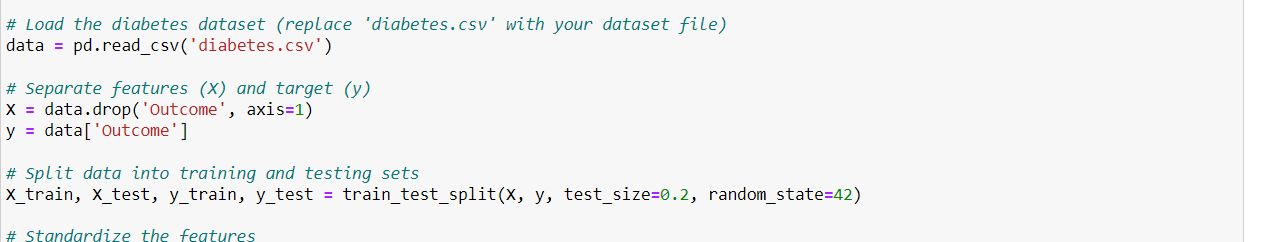
* **Pregnancies**
* **Glucose**
* **Blood pressure**
* **Skin Thickness**
* **Insulin**
* **BMI**
* **Diabetes Pedigree Function**
* **Age**
* **Outcome**

**EX:**

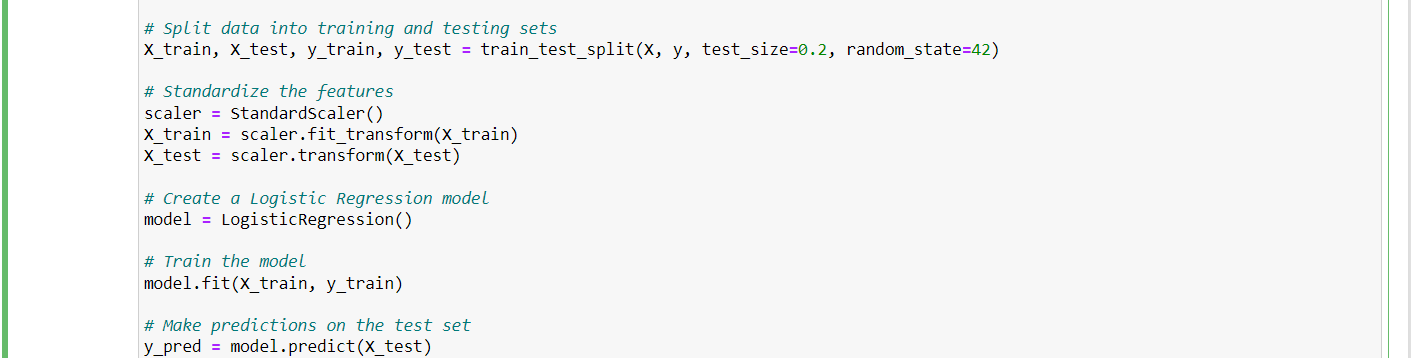
1. **Data preprocessing:**

**It is used to handle missing values, feature selection, and feature engineering**

**By**

****

1. **Split the data into training and testing sets:**

****

1. **Evaluating Model Performance:**

**By evaluating the performance of the model by the following parameters**

* **Accuracy**
* **Precision**
* **F1\_score**
* **Recall**
* **ROC AUC: The ROC AUC (Receiver Operating Characteristic Area Under the Curve)**
* **Confusion Matrix**

