**Name :Rajnish Prasad Kalwar USN: 1sv21ad026 Branch:AI&DS**

**Project -1: Personal Website URL (Github)**

**Rajnishkalwar.github.io**

**Project -2: Docker Image URL (DockerHub)**

**rajnishkalwar123/rajnishkalwar\_1sv21ad026**

**Project -3: Data Science Project Details**

**Title of the Project:**

**Diabetes Prediction**

**Project Description:**

**What is Diabetes?**

**Diabetes is a chronic disease that occurs when the pancreas is no longer able to make insulin, or when the body cannot make good use of the insulin it produces. Learning how to use Machine Learning can help us predict Diabetes. Let’s get started!**

**About this project :-**

**The objective of this project is to classify whether someone has diabetes or not.**

**Dataset consists of several Medical Variables(Independent) and one Outcome Variable(Dependent)**

**The independent variables in this data set are :-'Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin','BMI', 'DiabetesPedigreeFunction', 'Age'**

**The outcome variable value is either 1 or 0 indicating whether a person has diabetes(1) or not(0).**

**About the Dataset**

**Pregnancies :- Number of times a woman has been pregnant**

**Glucose :- Plasma Glucose concentration of 2 hours in an oral glucose tolerance test**

**BloodPressure :- Diastollic Blood Pressure (mm hg)**

**SkinThickness :- Triceps skin fold thickness(mm)**

**Insulin :- 2 hour serum insulin(mu U/ml)**

**BMI :- Body Mass Index ((weight in kg/height in m)^2)**

**Age :- Age(years)**

**DiabetesPedigreeFunction :-scores likelihood of diabetes based on family history)**

**Outcome :- 0(doesn't have diabetes) or 1 (has diabetes)**

**Algorithm Used:**

**1. K Nearest Neighbours :-**

**KNN algorithm, is a non-parametric algorithm that classifies data points based on their proximity and association to other available data.**

**2. Naive Bayes :-**

**Naive Bayes is classification approach that adopts the principle of class conditional independence from the Bayes Theorem. This means that the presence of one feature does not impact the presence of another in the probability of a given outcome, and each predictor has an equal effect on that result.**

**3. Support Vector Machine :-**

**It is typically leveraged for classification problems, constructing a hyperplane where the distance between two classes of data points is at its maximum. This hyperplane is known as the decision boundary, separating the classes of data points (e.g., has diabetes vs doesn't have diabetes ) on either side of the plane.**

**4. Decision Tree**

**5. Random Forest**

**6. Logistic Regression**

**Flow Diagram:**

**1. Import Required Libraries**

**2. Loading the dataset**

**3. Exploratory Data Analysis**

**4. Data Visualization**

**5. Feature Selection**

**6. Handling Outliers**

**5. Split the Data Frame into X and y**

**7. TRAIN TEST SPLIT**

**Results Screenshots:**



