Project Proposal: Diabetes Detector Using Machine Learning

1. Project Idea:
This project, Diabetes Detector, aims to develop an intelligent system capable of predicting an individual's likelihood of having diabetes based on health parameters. The system will use machine learning methods to analyze simple health indicators, such as glucose levels, blood pressure, BMI, and age, to make predictions. This project addresses the need for early detection, enabling timely intervention and reducing complications associated with diabetes.

2. Relevance to Sustainable Development Goals (SDGs):
This project aligns with the following SDGs:

- **SDG 3: Good Health and Well-being – Early diagnosis of diabetes helps reduce mortality and morbidity associated with complications like heart disease and kidney failure.**

- **SDG 9: Industry, Innovation, and Infrastructure – The use of AI and machine learning provides innovative and cost-effective solutions for early disease detection.**

3. Literature Examples:

- **Example 1:** Diabetes Prediction Using Machine Learning Techniques" by Patel et al. (2020) highlights the effectiveness of using machine learning algorithms like Logistic Regression and Decision Trees to predict diabetes.

- **Example 2:** Pima Indians Diabetes Dataset Analysis" by Smith et al. (2019) focuses on using the Pima dataset for predicting diabetes risk based on diagnostic health measurements**.**

These studies demonstrate the success of AI in identifying diabetes risk, which inspires this project's approach.

4. Describe Your Data:
The primary dataset is the Pima Indians Diabetes Database, originally sourced from the National Institute of Diabetes and Digestive and Kidney Diseases and accessed via Kaggle. The dataset contains key health indicators such as:

- **Features:** Pregnancies, Glucose, Blood Pressure, BMI, Age, Insulin, Skin Thickness, and Diabetes Pedigree Function.

- **Format: CSV file, approximately 768 records.**

- **Preprocessing Steps**: Handling missing values, normalization, and feature scaling.

5. Approach (Machine Learning or Deep Learning):
We will adopt a Machine Learning approach due to the structured and tabular nature of the data. Algorithms like Logistic Regression, Random Forest, and Support Vector Machines (SVM) will be evaluated for their accuracy in predicting diabetes. Machine learning is ideal for this task as it enables quick and accurate predictions while keeping the computational complexity low.