Dataset

Diabetes Detection: Machine Learning Classification Model for Predicting Diabetes

1. Data Source & Data Link

The dataset used for this project is the Pima Indians Diabetes Database, which is widely utilized for diabetes prediction models.

- **Data Source**: Kaggle
- **Data Link**: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

2. Libraries Required for Project

The following libraries are required for the development and deployment of the diabetes prediction model:

1. Pandas: Used for data manipulation and analysis, allowing for efficient handling of the dataset.

2. NumPy: Essential for numerical computations and array operations.

3. Matplotlib and Seaborn: Used for data visualization, helping to understand data distributions and relationships.

4. Scikit-learn: The main machine learning library used for building, training, and evaluating the classification model. It includes tools for data preprocessing, model selection, and performance evaluation.

5. Streamlit: Utilized for designing and deploying the web application, providing an interactive and user-friendly interface for the model.

6. Jupyter Notebook: Provides an interactive environment for code development and experimentation.