Diabetes Prediction

Introduction :

- Diabetes is a long-term disease that affects millions of people.
- It happens when the body cannot properly use sugar (glucose).
- Detecting diabetes early can help people take steps to stay healthy.
- The Diabetes Prediction tool uses artificial intelligence to analyze health data and predict the chances of having diabetes.
- This report explains how the tool works, what libraries it uses, and how it processes data.

Libraries :

This project uses the following Python libraries:

- **pandas** - Helps manage and process data in table format.
- **tensorflow.keras** - Helps build and train artificial intelligence models.
- **sklearn (scikit-learn)** - Helps prepare data for the model and split it for training.
- **Streamlit** - Helps create a simple web interface for users.

Process Overview :

1. Data Loading and Preprocessing

- The dataset comes from `diabetes.csv`, which contains health records of people, including glucose levels, BMI, blood pressure, and insulin levels.
- The data is divided into **features (X)** and **target variable (y)**, where `y` tells if a person has diabetes (1) or not (0).
- The data is standardized using `StandardScaler()` so that all values are on a similar scale, improving model accuracy.

2. Explanation of Features

Each record in the dataset contains the following health information:

1. **Pregnancies** - The number of times a woman has been pregnant. More pregnancies may increase the risk of diabetes during pregnancy (gestational diabetes).
2. **Glucose** - The amount of sugar in the blood. High levels suggest a risk of diabetes.
3. **Blood Pressure** - The force of blood against the arteries. High blood pressure is linked to diabetes and heart disease.
4. **Skin Thickness** - The thickness of skin on the arm, measured in mm. This helps estimate body fat.
5. **Insulin** - The level of insulin in the blood. Low or high insulin levels can indicate diabetes.
6. **BMI (Body Mass Index)** - A number calculated using weight and height. A higher BMI means more body fat, increasing the risk of diabetes.
7. **Diabetes Pedigree Function** - A number that estimates the chance of getting diabetes based on family history.
8. **Age** - Older people have a higher risk of getting diabetes, especially type 2 diabetes.

3. Model Training and Prediction

- The dataset is divided into **training data (80%)** and **testing data (20%)** using `train_test_split()`.
- A deep learning model is loaded from `ann_model.keras`. The model includes:
  - **Dense layers** - To learn patterns in the data.
  - **Dropout layers** - To prevent overfitting.
  - **Activation functions** - To improve predictions.
- The trained model predicts whether a user is diabetic or not based on their health data.

4. User Interface

The **Streamlit-based UI** allows users to enter their health details, such as:

- **Number of pregnancies**
- **Glucose level**
- **Blood Pressure**
- **Skin Thickness**
- **Insulin Levels**
- **BMI (Body Mass Index)**
- **Diabetes Pedigree Score**
- **Age**

After entering the data, the model gives a result:

- **"You are healthy"** if the diabetes risk is low.
- **"You are Diabetic"** if the risk is high.

5. Prediction Results and Visualization

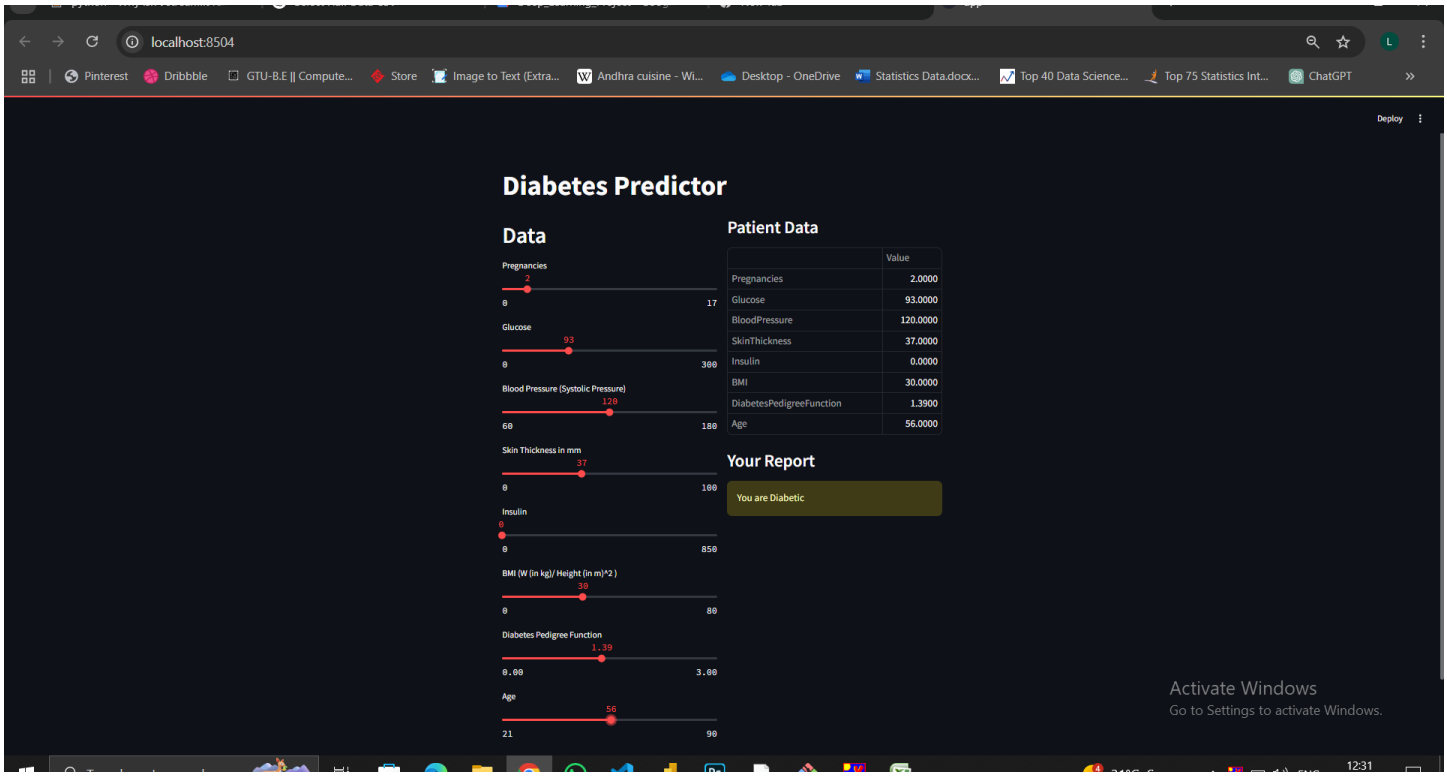
- The user inputs are displayed in a structured format.
- The model prediction is shown using **success/warning messages**.
- The results help users understand their health risks in real-time.

Conclusion :

The **Diabetes Prediction** tool helps analyze health data and provides a basic check for diabetes risk. By using artificial intelligence and a user-friendly interface, it helps users become aware of their health. Future improvements can include adding more real-world medical data and making the model even more accurate.

Output Images :

1.



2.