# Diabetes Prediction System 🩺

A web-based application built with **Streamlit** to predict the likelihood of a patient having diabetes using a machine learning model.

## 🎯 About the Project

This application provides a user-friendly interface for healthcare professionals or individuals to input patient data and receive an instant prediction on their diabetes status. The model is trained on the **Pima Indians Diabetes Dataset**, which contains medical diagnostic measurements for female patients.

The Random Forest model typically achieves an accuracy of approximately **75-80%** on this dataset.

## ✨ Features

* **Interactive Sidebar:** Allows users to input various health metrics such as Pregnancies, Glucose, Blood Pressure, BMI, etc.
* **Real-time Prediction:** Provides an immediate "Diabetic" or "Not Diabetic" result based on the user's input.
* **Visualized Patient Report:** Generates comparative static scatter plots for each metric, showing where the user's data point falls relative to the healthy and unhealthy groups in the training data. These plots are generated using **Matplotlib** and **Seaborn**.
* **Model Accuracy Display:** Shows the accuracy of the trained Random Forest Classifier on the test data.

## 🚀 Installation

### Prerequisites

Ensure you have Python installed on your system.

### Steps

1. **Clone the repository:**  
   git clone https://github.com/your-username/your-repository-name.git  
   cd your-repository-name
2. **Create and activate a virtual environment (recommended):**  
   python -m venv venv  
   # On Windows  
   venv\Scripts\activate  
   # On macOS/Linux  
   source venv/bin/activate
3. **Install the required dependencies:**  
   pip install -r requirements.txt

## 👨‍💻 Usage

To run the application, execute the following command in your terminal:

streamlit run app.py

The application will open in your default web browser at http://localhost:8501.

## 📂 File Structure

* app.py: The main Streamlit application script containing the UI, model training, and prediction logic.
* diabetes.csv: The dataset used to train the machine learning model.
* requirements.txt: A list of Python libraries and their versions required to run the application.
* Procfile: Configuration for deployment on platforms like Heroku.
* setup.sh: A shell script for setting up the Streamlit configuration for deployment.

## 📦 Dependencies

The project uses the following libraries, all of which are listed in requirements.txt:

* streamlit
* pandas
* numpy
* matplotlib
* seaborn
* scikit-learn
* Pillow
* plotly

## 🧠 How It Works

The app.py script performs the following steps:

1. **Data Loading:** Loads the diabetes.csv dataset.
2. **Data Preparation:** Splits the data into features (X) and target (y), and then into training and testing sets.
3. **Model Training:** A RandomForestClassifier is trained on the training data. **Note:** The model is trained dynamically each time the app.py script is run.
4. **Prediction:** The trained model predicts the outcome based on the user's input.
5. **Visualization & Reporting:** The application displays the prediction result and interactive graphs to visualize the patient's data.

## 📄 License

This project is licensed under the MIT License. See the LICENSE file for details.