Diabetes Prediction Tool: Simple and Interactive Application

We created a user interactive app that helps predict the risk of diabetes and visualize data based on the values given by the user. Using **machine learning** and **Streamlit**, this tool makes it easy for users to explore diabetes-related data and get predictions.

1. model_pkl.py

- We used a machine learning algorithm called **GradientBoostingClassifier** because it gave us accurate results for the diabetes dataset.
- This program saves the trained model in a file called Diabetes_Prediction_Model.pkl. This file helps us make predictions quickly without needing to retrain the model every time.

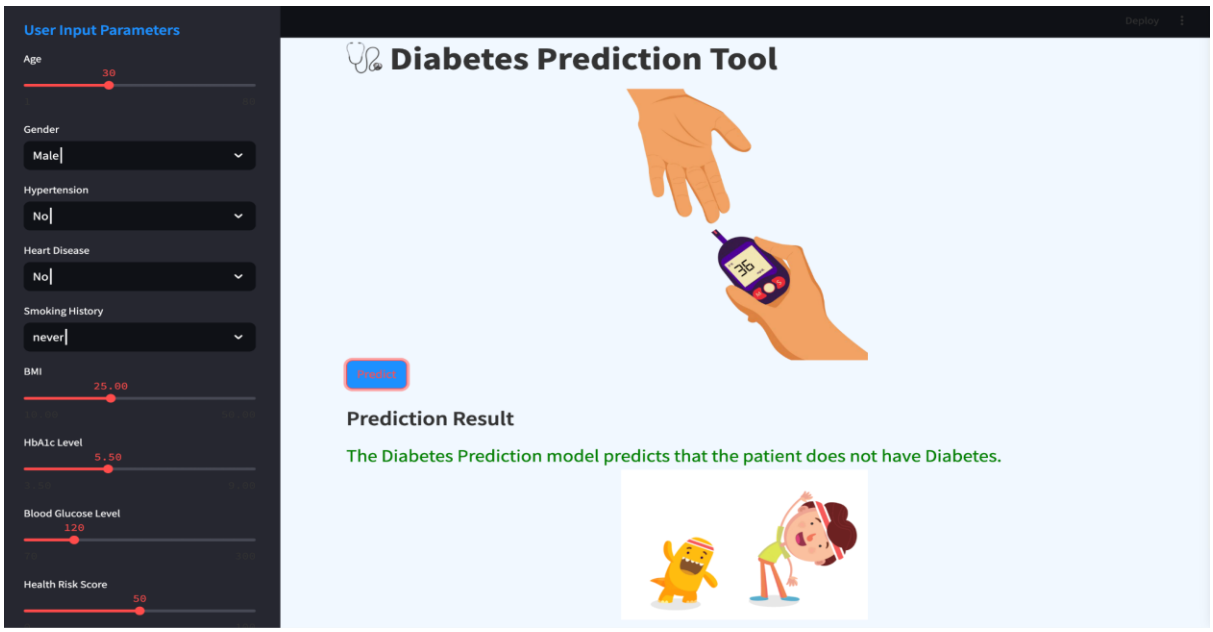
2. app.py

- This is the main application built with **Streamlit**, where users interact with the tool by changing the values of features on the side bar.
- The model is used here to predict the risk of diabetes based on the information users provide.
- The app includes an easy-to-use interface where users can adjust the feature values and see prediction results instantly.

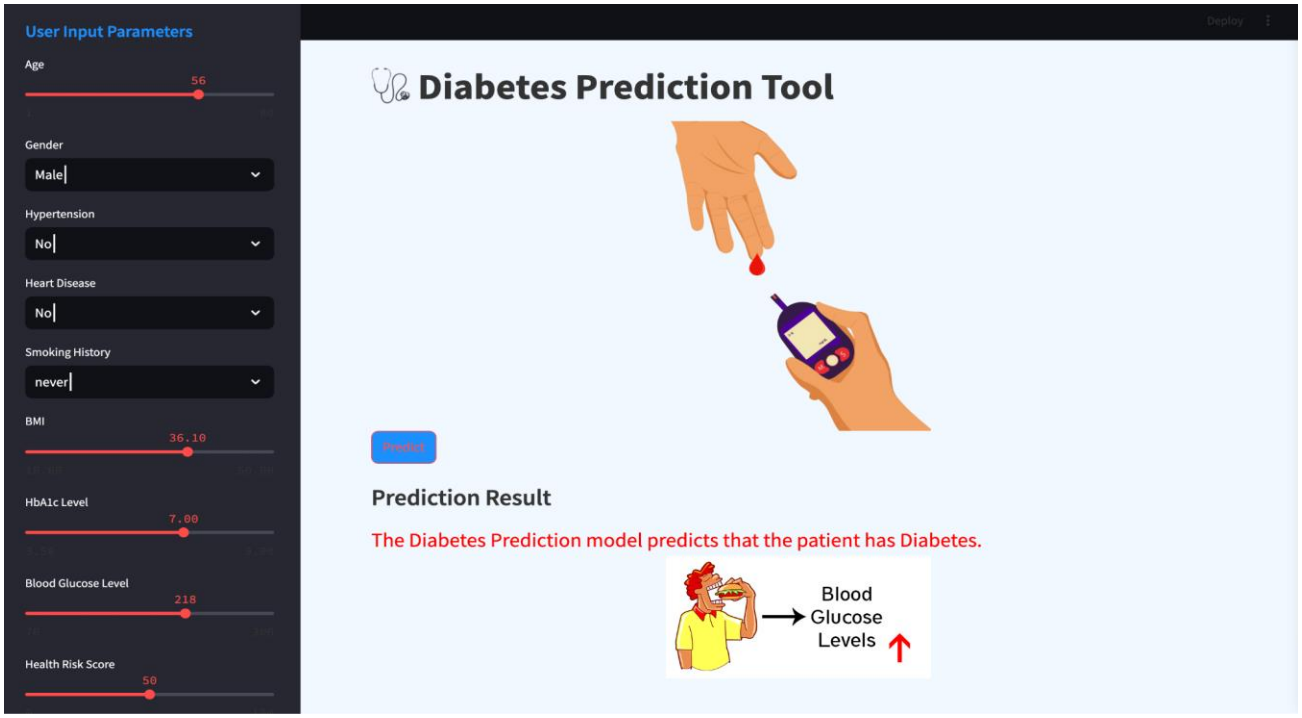
What the App Offers

- **Accurate Predictions:** We used Gradient Boosting, a reliable algorithm, to ensure high accuracy.
- **Interactive Data Visualizations:**
  - Scatter plots to show relationships between different data points.
  - Heatmaps to see how features are connected.
  - Distribution comparisons to easily spot trends.
- **Simple and Intuitive Design:** The app is designed so anyone, even without technical knowledge, can use it easily.

If the person doesn't have any diabetes, it will show the message similar to the below image.



Incase, if a person is found that they have diabetes based on the demographics and the hypertension, BMI, HBA1C level etc., the application will shows the user the prediction result like the below image.



visual.py: Visualizations and Insights

Our app helps users explore diabetes dataset and gain valuable insights through interactive visualizations. Here's what it offers:

Key Features

1. Dataset Overview

- Users can filter and explore the dataset easily.
- A preview of the dataset is displayed using `st.write()`, showing the rows and columns (dataset shape) for a quick overview.
- Users can download the filtered dataset for their own analysis.

2. Distribution Plots

- Interactive histograms with smooth curves (Kernel Density Estimation) let users see how data is distributed.
- Users can pick numeric columns to create custom visualizations.

3. Correlation Heatmap

- A heatmap shows how different numeric columns are related to each other, using color to highlight strong or weak connections.
- This makes it easier to spot patterns in the data.

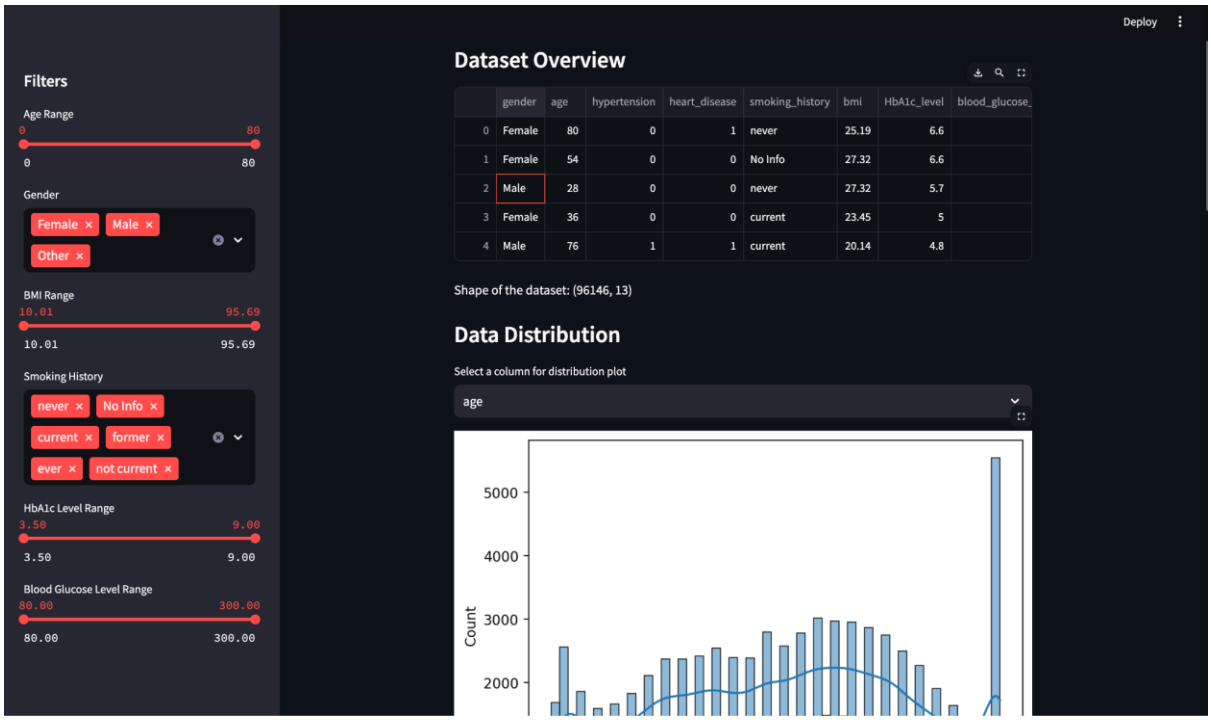
4. Scatter Plot

- Interactive scatter plots created with Plotly Express let users compare two numeric columns.
- Points are colored based on diabetes status, making it easy to see differences.

5. Box Plot

- Box plots show the range and spread of selected variables.
- They are grouped by diabetes status, helping users compare differences between people with and without diabetes.

This app makes exploring and understanding diabetes data simple, visual, and interactive!



diabetes_prediction_dataset.csv: Contains the dataset used for model training and visualization.

Running the Application

1. Execute model_pkl.py to generate the pickle model.
2. Using the pickle file generated from the model_pkl.py into the app.py file. We have to run python -m streamlit run app.py to predict whether patients have diabetes or not using the application.
3. To get the data insights of the dataset, run python -m streamlit run visual.py.

By providing a working user interface through Streamlit, we eliminated the need for users to interact with code directly. Allowed users to input their own data through form fields in the

GUI, making it easily understandable to any non-technical person. Offering immediate feedback on predictions and interactive data visualizations, enabling users to gain insights and manipulate data views as needed.

Diabetes_prediction_application : https://diabetes-prediction-ap.streamlit.app/

Diabetes_predication_dataset_visualisation: https://diabetes-prediction-visual.streamlit.app/