Diabetes Risk Predictor — User Guide

1. Introduction

This tool predicts the risk of diabetes based on key health indicators such as glucose level, BMI, age, blood pressure, and family history. It uses a decision tree model trained on the well-known Pima Indians Diabetes Dataset, which includes medical data from adult women of Pima Indian heritage. This dataset is widely used for teaching and research in machine learning. The goal of this tool is to provide a simple, interactive way to test the model by entering real-world-like health values and instantly see whether the predicted diabetes risk is high or low.

2. How to Run

Open MATLAB

Make sure the files diabetesPredictorApp.m and predictDiabetes.m are in the same folder

Run the command: diabetesPredictorApp

The app window will appear.

3. Input

Enter the following values into the app's input fields:

Input	Description	Units	Typical Range
Glucose	Blood glucose level	mg/dL	70 – 180
BMI	Body Mass Index	kg/m^2	15 – 40
Age	Age of the person	Years	20 – 80
Diabetes Pedigree Function	Family history/genetic risk factor	Unitless (0–2)	0 – 2
Pregnancies	Number of times pregnant	Count	0 – 15

Input	Description	Units	Typical Range
Blood Pressure	Diastolic blood pressure	mm Hg	50 – 100

4. Output

After entering all values, click **Predict**

The app will display: **Risk : High** (red text) or **Risk: Low** (green text)

5. Model Background

The prediction logic is based on a **decision tree** trained with the **J48 algorithm** (an implementation of the C4.5 algorithm) using the **WEKA** machine learning toolkit. The model was trained on the **Pima Indians Diabetes Dataset**, which is a well-known public dataset from the UCI Machine Learning Repository.

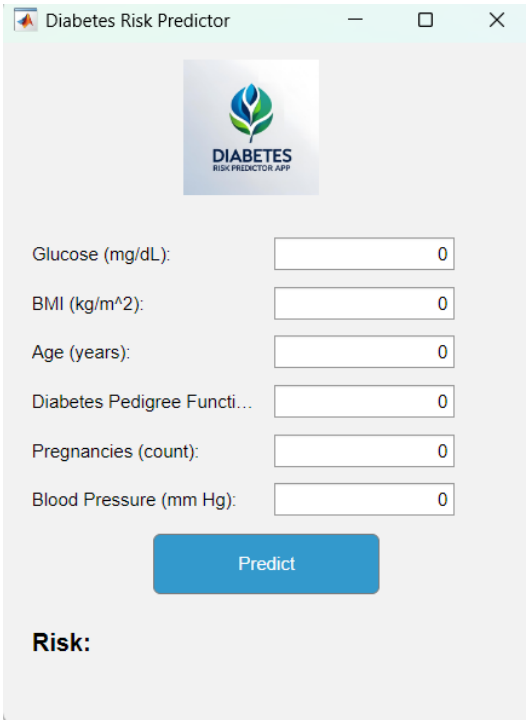
This dataset contains medical information about adult women of Pima Indian heritage, including attributes such as glucose level, BMI, age, number of pregnancies, blood pressure, and a diabetes pedigree function that indicates genetic risk. The target variable shows whether each patient was diagnosed with diabetes.

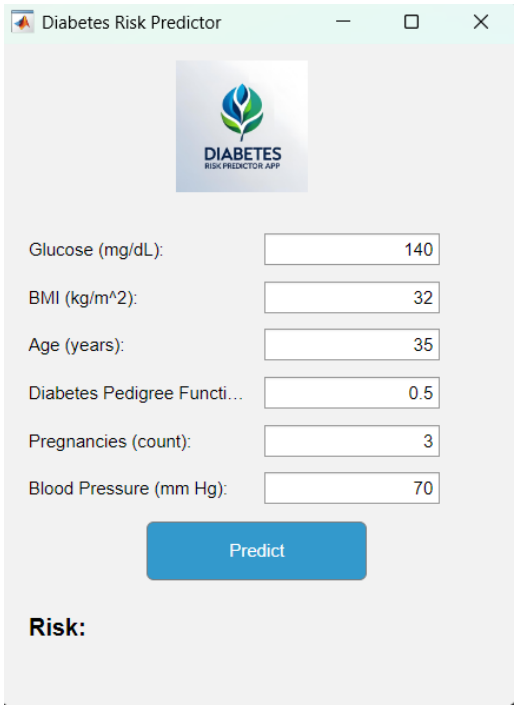
The J48 decision tree algorithm learns patterns in the data by splitting it into branches based on thresholds (e.g., glucose > 127). These splits result in a tree of if-else rules, which are implemented in the predictDiabetes.m file as MATLAB if statements.

This simple rule-based logic makes it easy to interpret how the prediction is made, which is why decision trees are commonly used for medical data exploration and teaching machine learning concepts.

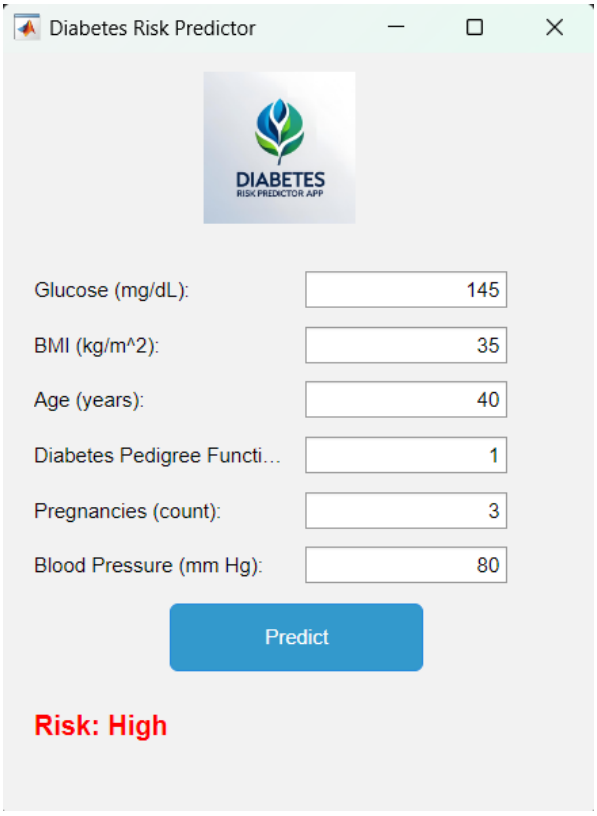
6. Screenshots

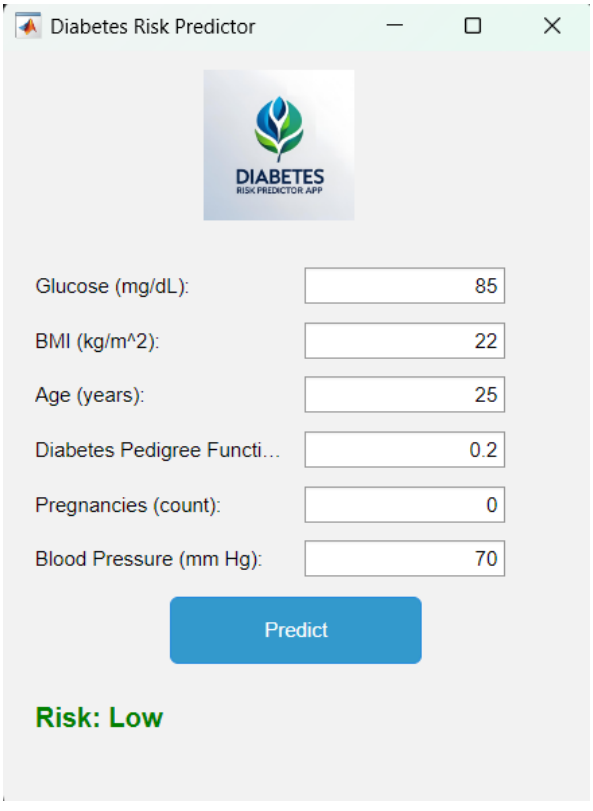
Main App Window (Full view) || Filled Inputs Before Prediction





Result Displayed





7. Troubleshooting & Tips

Ensure that all input fields are filled with valid numeric values; empty fields will cause an error. Negative values are not allowed for any input. The app validates input ranges as follows: Glucose should be between 50 and 300 mg/dL, BMI between 10 and 60 kg/m^2 (values outside this range may reduce prediction accuracy), Age between 0 and 120 years, Diabetes Pedigree Function (DPF) between 0 and 2 (unitless), Pregnancies must be a whole number between 0 and 20 and is intended for female patients, and Blood Pressure should be between 30 and 200 mm Hg. If you receive unexpected results, double-check that all inputs are entered using the correct units as specified in the input labels. The. Please make sure to enter whole numbers for Pregnancies as decimal values will cause an error. If the app displays an alert message, read it carefully and correct the input before trying again.